University of London
London School of Hygiene & Tropical Medicine

ESTIMATING THE IMPACT OF INFLUENZA VACCINATION AND ANTIGENIC DRIFT ON INFLUENZA-RELATED MORBIDITY AND MORTALITY IN ENGLAND & WALES USING HIDDEN MARKOV MODELS

Andrea Gail Mann

A Thesis submitted for the degree of Doctor of Philosophy

2010

Declaration

I, Andrea Gail Mann, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Influenza causes substantial morbidity and mortality in some influenza seasons, especially among the elderly. Influenza seasons dominated by circulation of influenza A/H3N2 virus tend to result in more morbidity and mortality than seasons dominated by influenza A/H1N1 or influenza B viruses. Influenza viruses undergo constant mutation, called antigenic drift, which is largely driven by host immunity. It has been shown that antigenic drift in influenza A/H3N2 virus proceeds in a punctuated, as opposed to continuous, fashion. A cluster of antigenically similar influenza A/H3N2 viruses appears to remain dominant for between 1 and 8 influenza seasons before being supplanted by a new cluster. Influenza seasons when a new cluster becomes dominant may result in higher morbidity and mortality than other seasons. Influenza vaccine effectiveness varies between influenza seasons because of the different subtypes in circulation and the degree of antigenic match between vaccine and circulating variants. In each influenza season in recent years, over 70% of the population of England & Wales aged ≥ 65 has been vaccinated, though the impact of this high coverage on population level morbidity and mortality is unknown. Multivariate time series models were fitted to reports of laboratory confirmed influenza, sentinel general practitioner (GP) consultations for influenza-like-illness, and all deaths registered to underlying pneumonia or influenza in England & Wales from 1975/76 to 2004/05. The models successfully distinguish influenza - attributable GP consultations and deaths from GP consultations and deaths that would be expected in the absence of influenza. This distinction is made jointly by the laboratory reports and the non-laboratory confirmed surveillance data. It is not possible to use the multivariate time series models to quantify the average effect of the appearance of a new cluster of influenza A/H3N2 virus variants, or vaccine impact, on influenza - attributable morbidity or mortality in the data analyzed. Reasons for this are discussed.

Contents

List of Tables

List of Figures

List of Abbreviations

A&E	Accident and Emergency
AC	All-cause mortality
AgD	Antigenic distance
AIC	Akaike Information Criterion
ARIMA	Autoregressive integrated moving average
BGR	Brooks-Gelman-Rubin
biIDR	Bivariate Poisson identity-link random effect
biLOGR	Bivariate Poisson log-link random effect
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
CrI	Credible interval
CT	Cluster transition
df	Degrees of freedom
EM	Expectation - Maximisation
exp	Exponentiated
GLM	Generalised linear model
GP	General practitioner
GPRD	General Practice Research Database
HA	Haemagglutinin

HI	Haemagglutination inhibition
HMM	Hidden Markov model
HPA	Health Protection Agency
HPA CfI	Health Protection Agency Centre for Infections
ICD	International Classification of Disease
IDR	Poisson identity-link random effect
ILI	Influenza-like-illness
LL	Log likelihood
LNR	Natural logarithm random effect
LOGR	Poisson log-link random effect
MAXRR	Maximum rate ratio
MC-error	Monte Carlo error
MCMC	Markov chain Monte Carlo
MSGP4	*Morbidity Statistics from General Practice: fourth national study 1991-92*
NA	Neuraminidase
NBIDR	Negative binomial identity-link random effect
NBLOGR	Negative binomial log-link random effect
NHS	National Health Service
ONS	The Office for National Statistics
OXMIS	Oxford Medical Information Systems
P&I	Pneumonia and influenza mortality
PPD	Posterior predictive density
RCGP	Royal College of General Practitioners

RNA	Ribonucleic acid
RR	Rate ratio
RSV	Respiratory syncytial virus
SD	Standard deviation
SES	Socioeconomic status
SIRS	Susceptible-Infected-Recovered-Susceptible
SQR	Square-root random effect
triIDR	Trivariate Poisson identity-link random effect
triLOGR	Trivariate Poisson log-link random effect
UIIP	Universal Influenza Immunisation Program
UK	United Kingdom
US	United States of America
VE	Vaccine effectiveness
VRU	Virus Reference Unit
WHO	World Health Organization
WRS	Weekly Returns Service

Acknowledgements

I would like to thank my supervisor, John Whittaker, for teaching me an enormous amount about Bayesian statistics and statistical methods for epidemiology and for providing gentle guidance and constructive comments. I think you did know how little I knew but you never made me feel dumb.

Thank you to my advisors, official and unofficial, for sharing of their knowledge and experience: Tony Ades, Nick Andrews, Ben Cooper, Alex Elliot, Paul Fine, Douglas Fleming, Nigel Gay, Punam Mangtani (associate supervisor), Aronrag Meeyai, Laura Rodrigues, Colin Russell, Derek Smith and Clarence Tam. Thank you to Al Ozonoff of the Boston School of Public Health for sending me the code, to fit a hidden Markov model in OpenBUGS, that got me started. Steve Whitbread provided invaluable computing assistance. I would also like to thank Sara Thomas for her friendship and for giving generously of her time when I'm sure she didn't have much of it. Thank you to Dan, Hopi, Paul and Maria for helpful comments on drafts.

Douglas Fleming, Alex Elliot and Michele Barley kindly donated Royal College of General Practitioners data extracts. Thank you to all of the physicians contributing to this surveillance scheme. Cleo Rooney and Emma Gordon from The Office for National Statistics provided mortality data. Laboratory data were supplied by Carol Joseph, Joy Field and Piers Mook of the Health Protection Agency Centre for Infections.

The National Institute for Health Research granted me a fellowship enabling me to undertake this work. Thank you to Laura Rodrigues for bringing their fellowship scheme to my attention.

I would like to acknowledge Dan, Clare and Alej. Your messages of support were beacons in the night.

Thank you, Mom, for your love, and Dad, for letting me know that you're proud of me.

Finally, all my love and thanks to Dan for doing everything else so I could do this.

1. Introduction

Influenza causes substantial morbidity and mortality in some influenza seasons, especially among the elderly. [1] Influenza-related morbidity or mortality is rarely laboratory confirmed. [4, 5] As such, indirect methods are employed in order to determine the relative impact of influenza seasons in terms of the morbidity and mortality caused by influenza. A limitation of the regression models most frequently used to estimate influenza - attributable morbidity and mortality is that they require the epidemiologist to delete some of the observed data and fit a model to the remaining data. This is to allow estimation of the expected morbidity or mortality that would have occurred had influenza not be circulating (e.g. [6, 7]). The observed morbidity or mortality that exceeds the model-predicted expected morbidity and mortality is deemed attributable to influenza. The various methods of deciding what data should be deleted from model fitting are more or less arbitrary, leaving the possibility of counting unrelated deaths or consultations as influenza - attributable (lowering the specificity of influenza - attributable mortality for influenza) or of missing truly influenza - attributable deaths or consultations (potentially biasing downwards estimates of the effect of determinants of influenza - attributable morbidity or mortality by obscuring the full extent of the variability in influenza - attributable morbidity and mortality between influenza seasons). An alternative method of estimating influenza - attributable morbidity or mortality is by regressing deaths or consultations on indicators of influenza virus circulation (e.g. [8, 9]). Regression on laboratory data may produce biased estimates because of long-term changes to the number of tests performed. [10] **Objective 1** of the work undertaken for the thesis was to develop a model in which the distinction of influenza - attributable from expected morbidity and mortality would be made by the model (not the epidemiologist). The specificity of non-laboratory morbidity and mortality data for influenza is increased by using both types of data (non-laboratory confirmed data and

laboratory data) to inform the distinction of influenza - attributable from expected rates of morbidity and mortality.

Objective 1: *to estimate the relative impact of influenza seasons in England & Wales between 1975/76 and 2004/05 in terms of GP consultations for influenza-like-illness (ILI) and deaths from pneumonia or influenza (P&I) by jointly modeling ILI, P&I and laboratory reports for influenza A virus using multivariate latent variable time series models.*

The latent variable models described in the thesis captured the relative impact of influenza seasons using a simple random effect mean shift (the ratio of average influenza - attributable to expected morbidity or mortality rates, by influenza season). The distinction of influenza - attributable from expected ILI or P&I was made by the model, informed by laboratory reports for influenza A. The laboratory data informed the timing, not the relative impact, of influenza seasons.

Influenza seasons dominated by circulation of influenza A/H3N2 virus tend to result in more morbidity and mortality than seasons dominated by influenza A/H1N1 or influenza B virus. [1] Influenza viruses undergo constant mutation, called antigenic drift, which is thought to be driven largely by host immunity. [11, 12] It has been shown that antigenic drift in influenza A/H3N2 virus occurs in a punctuated, as opposed to continuous, fashion. [13] A cluster of antigenically similar influenza A/H3N2 viruses appears to remain dominant for between 1 and 8 influenza seasons before being supplanted by a new cluster. Work by others suggests that influenza seasons when a new cluster becomes dominant may result in higher morbidity and mortality than other seasons. [14] The average effect of cluster transition seasons on influenza - attributable morbidity and mortality in England & Wales (or, indeed, globally) has not been quantified. The effect of cluster transition seasons can be estimated by expressing the chosen metric for relative impact of influenza seasons (the random effect mean shift) as dependent on a binary variable for cluster transition seasons. **Objective 2** of the work undertaken for the thesis was to use the models developed for objective 1 to estimate the average effect of a cluster transition on the mean

shift in P&I and ILI in each influenza season between 1975/76 and 2004/05 during which time there were 9 cluster transitions (in 1975/76, 1977/78, 1979/80, 1987/88, 1989/90, 1992/93, 1995/96, 1997/98 and 2002/03).

***Objective 2**: to use the multivariate models developed for objective 1 to estimate the mean effect of cluster transitions in influenza A/H3N2 virus evolution on the mean shift in P&I and ILI by age group.*

The variability in the size of cluster transitions in terms of the degree of antigenic drift was allowed for in a supplementary analysis where the random effect mean shift was expressed as dependent on a quantitative variable for the size of cluster transitions.

Influenza vaccine efficacy against laboratory-confirmed influenza-like-illness in the elderly has been demonstrated. [15] Vaccine effectiveness varies because of influenza season-specific factors such as the relative impact of the season in terms of morbidity and mortality and the degree of antigenic match between vaccine and circulating variants. [16, 17] Between 1989/90 and 2004/05, yearly vaccine coverage of the $\geq$ 65 age group in England & Wales has increased from 24 to 71%. The impact of this change in coverage on population level morbidity and mortality is unknown. **Objective 3** of the work undertaken for the thesis was to estimate the impact of each unit increase in vaccine coverage of the elderly on morbidity and mortality by expressing the random effect mean shift as dependent on vaccine coverage.

***Objective 3**: to estimate the impact on the mean shift in ILI and P&I for the $\geq$ 65 age group, and for the other age groups, per unit increase in yearly vaccine coverage of the $\geq$ 65 age group.*

The thesis has the following structure. Chapter 2 is a survey of the literature on methods to estimate the relative impact of influenza seasons in terms of morbidity and mortality. Evidence for an average inflating effect of cluster transitions on influenza - attributable morbidity and mortality is

reviewed, as is evidence for a population impact of vaccination of the elderly. In chapter 3, P&I and ILI data for England & Wales are described. Models used in the thesis - the frequentist generalised linear models used initially to estimate the shape of long-term trend in P&I and ILI and the Bayesian Markov chain Monte Carlo methods used to fit two-state latent variable time series models to P&I, ILI and laboratory reports for influenza A - are introduced. In chapter 4, the shape of the long-term trend in P&I and ILI by age group from 1970 to 2005 is presented. Crude associations between peak P&I and ILI in each influenza season, by age group, and the exposures of interest (cluster transitions and vaccine coverage of those ≥ 65) are explored. Chapters 5 and 6 describe development of univariate and multivariate two-state hidden Markov models (latent variable time series models) used to estimate the relative impact of the influenza seasons between 1975/76 and 2004/05 in terms of P&I and ILI. In chapter 7, crude associations between the mean shift in P&I and ILI, by age group, and exposures of interest are examined. Results from models including a dependency between the mean shift and cluster transitions or vaccine coverage of the ≥ 65 age group are presented. Also in this chapter, the relative impact of influenza seasons in England & Wales estimated using other methods are compared to estimates using the multivariate latent variable time series models fitted in the thesis. Findings related to the average effect of cluster transitions and vaccine impact are also placed in context of what is already known. Chapter 8 summarises the main findings and suggests directions to take for the future.

2. Literature Review

2.1. Introduction

This chapter is structured as follows. Background on influenza virus and disease is provided in section 2.2. The variability in estimates of influenza - attributable morbidity and mortality between influenza seasons and between studies is introduced in section 2.3. In section 2.4 the methods that have been used by others to estimate influenza - attributable morbidity and mortality are critically reviewed. Section 2.5 introduces key potential determinants of the variability in influenza - attributable morbidity and mortality between influenza seasons, including the exposures of interest in the work described in the thesis: large antigenic drift events in influenza A/H3N2 virus evolution and vaccination. Sections 2.6 and 2.7 review the plausibility of and evidence for an effect of large antigenic drift events, defined as antigenic cluster transitions by Smith *et al.* [13], on mean influenza - attributable morbidity and mortality in influenza seasons. The plausibility of and evidence for an impact of increasing yearly vaccine coverage of the $\geq$ 65 age group on mean influenza - attributable morbidity and mortality across age groups are reviewed in sections 2.8 and 2.9. Finally, the rationale for the work described in the thesis is given in section 2.10.

2.2. Background

Influenza virus is a member of the Orthomyxoviridae family. There are three types of influenza viruses - A, B, and C - though only A and B cause widespread outbreaks in humans. [18] Influenza virus is transmitted via large droplets (expelled during coughing and sneezing), aerosols (tiny droplets) and fomites. [18,19]

Influenza virus has a segmented ribonucleic acid (RNA) negative sense single-stranded genome. The 8 genome segments encode 10 gene prod-

ucts: PB1, PB2, and PA polymerases, haemagglutinin (HA), neuraminidase (NA), NP, M1, and M2 proteins, non-structural protein 1 and non-structural protein 2. [20] HA has two subunits, HA1 and HA2, and is critical for pathogenesis. [21] The HA is one of 2 major antigenic determinants recognised by host neutralising antibodies (the other is NA). The HA is involved in attachment to the host cell, via sialic acid on the host cell surface, and cell entry. [21] The NA is essential to the release of new virus particles from infected host cells via cleavage of glycosidic linkages to sialic acid residues binding new virus particles to the host cell surface. [21] The NA prevents viral aggregation and facilitates viral dispersion in mucus. It is an important target for antivirals like oseltamivir and other neuraminidase inhibitors. [18] The M2 protein forms an ion channel (that is blocked by the antiviral amantadine) which regulates the pH of the virus and enables early viral replication. [18] Random assembly of the 8 different RNAs into new virions leads to progeny viruses with new combinations of genes (reassortment) when a host cell is infected by two different virus variants. [20] The NA, HA and M2 proteins are embedded in the envelope on the surface of the virus. Subtypes of influenza A virus are defined by their HA and NA (e.g. H3N2, H1N1); different subtypes of influenza B virus have not been identified.

Within subtypes of influenza A virus there are variants which differ from one another genetically and antigenically, that is in the degree to which they elicit an antibody response in the host. Antigenic variants arise due to a process of mutation called antigenic drift. [18] Antigenic drift is the stepwise accumulation of mutations in the HA and NA which means that antibodies raised against a previous variant (in the evolutionary line) will be progressively less able to recognise, and neutralise, newer variants. Antigenic drift results in hosts becoming susceptible to influenza virus infection anew. Antigenically drifted variants also arise through reassortment. The high rate of mutation in influenza virus even compared to other single stranded RNA viruses is due to relatively low fidelity of the RNA polymerase and no error-checking capability. [21] Antigenic drift has been shown to happen more in influenza A than B viruses. [22]

In contrast to antigenic drift, antigenic shift is the emergence or reemergence of influenza A viruses against which the population has virtually no immunity; antigenic shift may lead to influenza pandemics if other conditions, such as good human to human transmission of the pandemic virus,

are met. Pandemics are sometimes characterised by high mortality because of the lack of immunity in the population. A shift in the age distribution of mortality from older people, who die from seasonal influenza, to younger people can occur during pandemics because of antigen recycling whereby similar HA or NA circulated long ago and only the oldest people have any immunity. [23–25] Influenza B virus has no animal reservoir so does not experience antigenic shift. [21]

Historically antigenic shift has happened when a virus with a new HA (with or without other new gene segments like a new NA) has infected humans. The new HA may have arisen directly from an avian reservoir [26] or via reassortment between human and avian viruses, sometimes via the mixing vessel of pigs. [27] The current pandemic H1N1 2009 virus is sufficiently diverged evolutionarily from seasonal influenza H1N1 viruses to mean most age groups are essentially fully susceptible. [28] The nomenclature used to define antigenic shift viruses therefore needs to be revised to reflect that antigenic shift can occur without a new NA or HA. [29] In the 20th Century there were 3 formally recognised pandemics - 1918 (H1N1), 1957 (H2N2) and 1968 (H3N2) - and 1 pandemic-like episode in 1977 when an H1N1 virus identical to viruses circulating in 1950 reemerged, [30] possibly as an accidental release from a laboratory. [31] Seasonal or inter-pandemic influenza viruses in circulation since the last pandemic are influenza A/H3N2 virus, influenza B virus and, since 1977, influenza A/H1N1 virus; a reassorted influenza A/H1N2 virus circulated for a short time starting in 2001. [32]

2.2.1. Key definitions

The period during which influenza virus is circulating in the community, the 'influenza season', is informed by routine [10] and sentinel laboratory surveillance. [33, 34] In the temperate Northern hemisphere influenza incidence displays pronounced seasonality; influenza virus typically circulates in the community between November and June with little recognised circulation in the summer months. [35] Morbidity and mortality from influenza related causes is also highly seasonal (e.g. figure 2.1). It should be noted that little laboratory testing for influenza virus is done in summer leading to potential underestimation of summer circulation of influenza virus. Influenza years will be defined as the first week of July to the last week of June.

Henceforth, the term 'normal' incidence will refer to the roughly sinusoidal seasonal pattern of respiratory morbidity and mortality and 'aberrant' to incidence in excess of 'normal' incidence. 'Aberrant' incidence is observed during most influenza seasons, but is of variable intensity. Epidemics are periods when 'aberrant' incidence is unusually high.

2.3. Relative impact of influenza seasons

In this section, the variability in estimates of influenza - attributable morbidity and mortality between influenza seasons and between studies is introduced. Around 5% of adults and 20% of children, globally, have symptomatic influenza A or B each year. [37] Influenza virus infection in humans causes both upper and lower respiratory symptoms. [18] Although infection is most common in the young, [38–40] morbidity and mortality occurs largely in the elderly [8,41–43] and in people with underlying cardiorespiratory disease or diabetes. [44] Deaths in the elderly are thought to be caused by secondary bacterial pneumonia. [45] Those who die as a result of influenza are often elderly people with comorbid respiratory or cardiovascular disease who are admitted to hospital from the community. [46]

Most diagnoses of influenza are not laboratory-confirmed. [4, 5] Most deaths associated with influenza do not have influenza mentioned on death certificates. [47] Morbidity and mortality attributable to influenza is typically ascertained through calculating numbers or rates of general practitioner (GP) consultations, for diagnoses such as influenza-like-illness (ILI), otitis media, acute bronchitis and asthma, [1–3] hospitalisations for respiratory disease, [1,9,48–50] and deaths from all-causes, [1,8,41,51,52] respiratory (pneumonia + influenza + bronchitis) and circulatory diseases, [8, 9,51–53] or pneumonia and influenza [7,54] in excess of the number or rate expected in the absence of influenza virus. Between 25 and 50% of people who consult a GP for ILI do so during outbreaks of influenza. [36,55] Excess mortality [1,56] and hospitalisations [1,48] during epidemics suggests at least a portion of these are caused by the influenza virus.

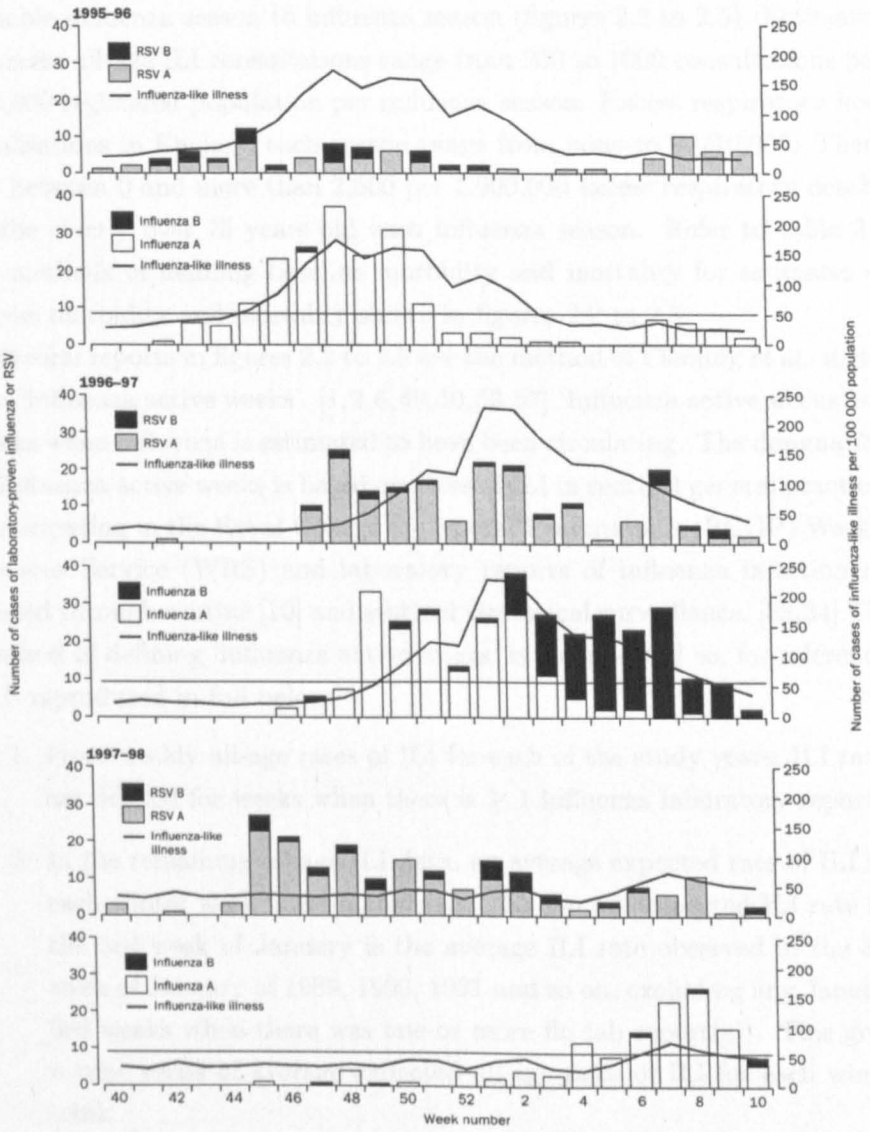


Figure 2.1.: **Weekly rates of general practitioner consultations for influenza-like-illness and numbers of virus isolates for influenza A, B and respiratory syncytial virus: 1995/96 - 1997/98.** Reproduced from [36]

2.3.1. In England & Wales

The burden of morbidity and mortality attributable to influenza is highly variable influenza season to influenza season (figures 2.2 to 2.5). Estimates of excess all-age ILI consultations range from 200 to 1600 consultations per 100,000 registered population per influenza season. Excess respiratory hospitalisations in England each season range from none to 70/10,000. There are between 0 and more than 2,500 per 1,000,000 excess respiratory deaths in the elderly over 75 years old each influenza season. Refer to table 2.1 for methods of defining baseline morbidity and mortality for estimates of excess morbidity and mortality shown in figures 2.2 to 2.5.

Several reports in figures 2.2 to 2.5 use the method of Fleming *et al.* to define 'influenza active weeks'. [1,2,6,49,50,53,57] 'Influenza active weeks' are weeks when influenza is estimated to have been circulating. The designation of influenza active weeks is based on rates of ILI in sentinel general practices participating in the Royal College of General Practitioners (RCGP) Weekly Returns Service (WRS) and laboratory reports of influenza infection reported through routine [10] and sentinel virological surveillance. [33,34] The method of defining 'influenza active weeks' is complicated so, for reference, it is reproduced in full below:

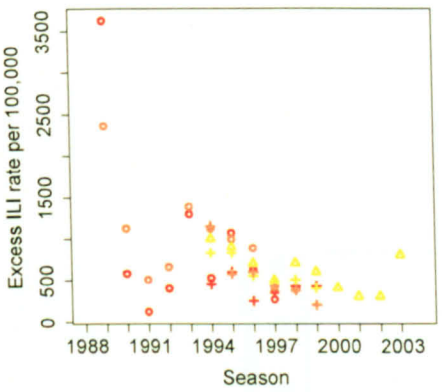
1. From weekly all-age rates of ILI for each of the study years, ILI rates are deleted for weeks when there is ≥ 1 influenza laboratory report.

2. In the remaining all-age ILI data, an average expected rate of ILI for each winter week is estimated (e.g. the average expected ILI rate for the 3rd week of January is the average ILI rate observed in the 3rd week of January of 1989, 1990, 1991 and so on, excluding any January 3rd weeks when there was one or more flu lab report(s)). This gives a time series of average expected all-age rates of ILI for each winter week.

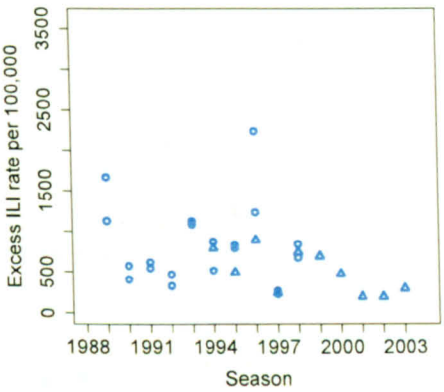
3. A 95% confidence interval (CI) is calculated around the time series of average expected all-age rates of ILI for each winter week. The upper 95% confidence limit on the time series of average expected rates of ILI is defined as the 'epidemic threshold'.

4. Observed all-age ILI rates for every week of the study years are compared with the epidemic threshold for that week. When the observed

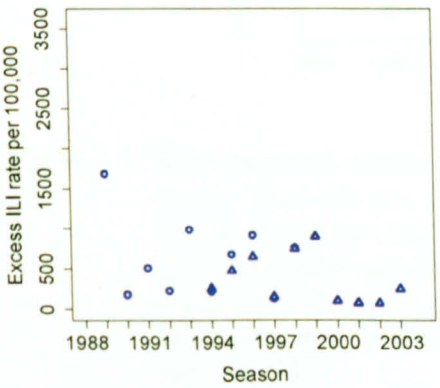
ILI rate exceeds the epidemic threshold, the week is defined as an 'influenza active week'.



(a) Children $\leq$ 15 yrs



(b) 15-64



(c) $\geq$ 65

Figure 2.2.: **Estimates of excess ILI in England & Wales.** In (a): red in the <1 yr age group, yellow is the 1-4 age group, pink is the 5-14 age group. Circles [1], '+' [2], diamonds [3].

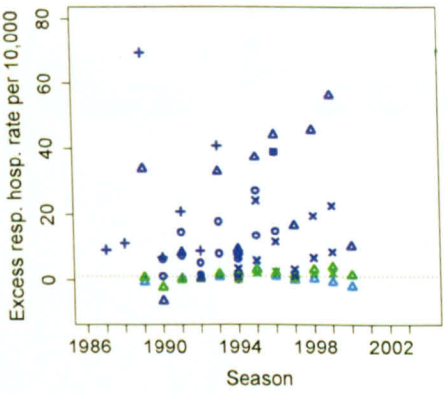
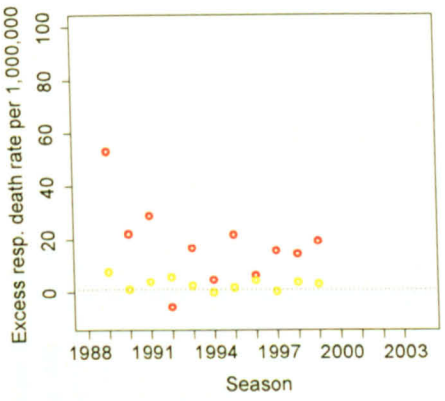
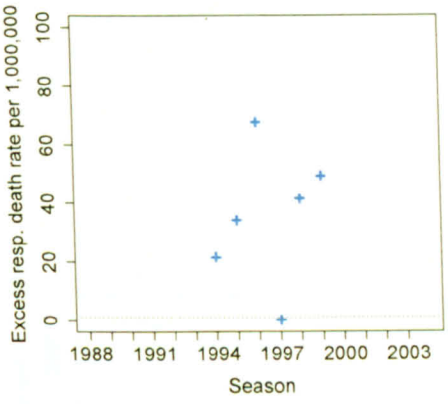


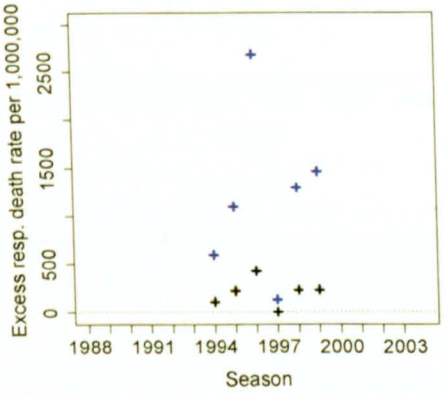
Figure 2.3.: **Estimates of excess respiratory hospitalisations in England.** Blue <25 yrs, green $\geq$ 25 and $\leq$ 65, purple $\geq$ 65. Circles [1] (when two circles are visible for the same influenza season, the higher estimate is the $\geq$ 75 age group, the lower estimate is the 65-74 age group), '+' pneumonia or influenza hospitalisations from selected health authorities [49], diamonds [50] and 'x' [53].



(a) Children $\leq$ 5 yrs

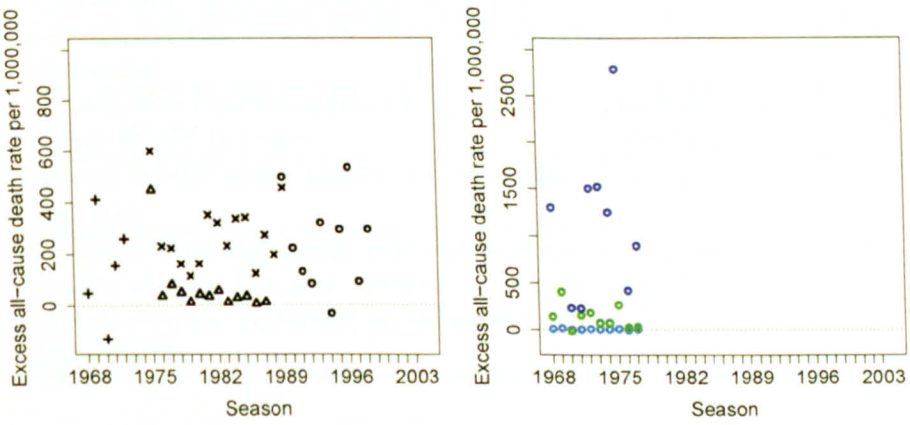


(b) 45-64



(c) $\geq$ 65

Figure 2.4.: **Estimates of excess respiratory mortality (pneumonia, influenza or bronchitis) in England & Wales.** In (a): red is the $\leq$ 1 yr age group, yellow is the 1-4 age group, both from [57]. (b) is from [53]. In (c): black is the 65-74 age group, purple the $\geq$ 75 age group [53].



(a) All-age

(b) 15-39, 40-64, $\geq$ 65 yrs

Figure 2.5.: **Estimates of excess all-cause mortality in England & Wales.** In (a) circles [1], '+' [58], diamonds [47] and 'x' [59]. In (b) blue is the 15-39 age group, green the 40-64 age group, purple the $\geq$ 65 age group, all from [51]

Table 2.1.: **Key to methods used to estimate baseline morbidity and mortality in figures 2.2 to 2.5**

Definition of baseline	**Table reference**	**refs**
Mean rate by age group in winter weeks outside of 'influenza active weeks'	circles fig 2.2; all fig 2.3; '+' fig 2.4; black circles fig 2.5	[1, 49, 50, 53]
Mean rate by age group in winter weeks outside of 'influenza active weeks' and 'respiratory syncytial virus (RSV) active weeks'[1]; excess apportioned between influenza, RSV if overlapped	'+' in fig 2.2; circles fig 2.4	[2,57]
Average morbidity in winter weeks excluding weeks encompassing $\geq$ 70% of laboratory reports that season (influenza or RSV; excess apportioned between them if overlapped)	diamonds fig 2.2	[3]
Seasons without 'very high peaks' or 'sizeable outbreaks'	'+' fig 2.5	[58]
Months with 0 influenza laboratory reports or lowest RCGP ILI rate	coloured circles fig 2.5	[51]
Lowest RCGP upper respiratory tract infection rate	'x' fig 2.5	[59]
Intercept from regressing total excess winter deaths (deaths Dec-Mar divided by average of Aug-Nov and Apr-July) on influenza registered deaths	diamonds fig 2.5	[47]

Variability in excess ILI between influenza seasons is similar across age groups (figure 2.2). Estimates of excess ILI for later influenza seasons [3] were lower in general than for earlier years. [1]. This may be related to the different ways baseline ILI was defined in the later vs. earlier reports (see table 2.1).

Excess respiratory hospitalisations are variable between influenza seasons and are higher for older age groups (figure 2.3). Estimates of excess pneumonia and influenza hospitalisations [49] ('+' in figure 2.3) are higher than estimates of excess respiratory hospitalisations [1, 50] (all other symbols in figure 2.3) because pneumonia and influenza hospitalisations are more spe-

[1] 'RSV active weeks' were defined as weeks with at least 200 laboratory reports for RSV in children <1 and then following the same procedure as for 'influenza active weeks'

cific to influenza. Pneumonia and influenza hospitalisations would therefore be expected to occur at lower rates outside of the influenza season. For a given influenza season, estimates from [53] (purple 'x') are lower than estimates from [50] (purple diamonds) because averaging rates around Christmas holiday weeks done in [53] lowered the estimated excess respiratory hospitalisation rate. There is no long-term trend in excess hospitalisations.

Excess respiratory mortality is variable season to season and is much higher in the older than younger age groups (figure 2.4). A similar pattern is seen in rates of excess all-cause mortality (figure 2.5). There is no clear long-term trend in excess mortality except in the youngest age groups. Estimates of excess all-age, all-cause mortality that rely on relative numbers of laboratory reports for influenza A and B in different influenza seasons [59] (black diamonds in figure 2.5) produce lower estimates, and less variability in excess mortality between influenza seasons, than other methods. This suggests that the relative number of laboratory reports between influenza seasons may not be a good proxy for the relative impact of influenza seasons.

A substantial and variable burden of influenza morbidity and mortality has been documented in many other settings. [60–67] Comparison of estimates between countries is difficult because of different diagnostic coding practices and because of different methods used to estimate morbidity and mortality in the absence of influenza virus.

2.4. Critique of methods to estimate influenza - attributable morbidity and mortality

In this section, six general methods to estimate influenza - attributable morbidity and mortality are critically reviewed. These general methods are Serfling-like least squares, ARIMA time series, linear or Poisson regression, rate difference, transmission dynamic and latent variable time series methods.

The cumulative influenza - attributable morbidity and mortality in influenza seasons has been estimated in many ways. Farr introduced the concept of 'excess mortality', attributable to influenza, over an expected mortality in the absence of influenza virus circulation. Farr subtracted the number of deaths that occurred during an influenza epidemic in 1847 from

the average monthly mortality rate to estimate mortality attributable to the epidemic. [68] Collins defined expected mortality as the median weekly number of deaths during non-epidemic years and excess as observed minus expected deaths. [69]

2.4.1. Serfling-like least squares

Serfling introduced the concept of fitting a sinusoidal curve by least squares to mortality data with epidemic weeks deleted. [70] This idea is based on the assumption that the winter increase in mortality is only partly attributable to influenza. Serfling expressed 4-weekly P&I, with 'aberrant' or 'epidemic' weeks deleted, as dependent on an intercept, linear trend and one Fourier term (one sine plus one cosine term) to model seasonality with a period of one year according to the formula

$$Y_t = \alpha + \beta t + \beta_1 \sin\frac{2\pi t}{12} + \beta_2 \cos\frac{2\pi t}{12}$$

where Y_t is the monthly mortality rate, α is the intercept, β is a linear term for long-term trend. β_1 and β_2 are coefficients for seasonality. Five influenza seasons worth of non-epidemic P&I data were used to predict the expected P&I for the sixth influenza season. Coefficients were estimated by least squares. Serfling then defined the 'epidemic threshold' for the sixth influenza season as 1.65 standard deviations (SD) above the predicted mortality for that season (figure 2.6). Excess P&I for the whole of the sixth influenza season was the sum of observed mortality that was greater than the epidemic threshold. This method has been widely used (e.g. [56, 58]). The Health Protection Agency Centre for Infections (HPA CfI) uses a Serfling-type approach to monitor excess mortality in England & Wales each winter. [71] The Centers for Disease Control and Prevention has used a Serfling-type model, modified by Lui and Kendal, fitted to past P&I incidence from 122 US cities to predict expected P&I for the following influenza season for many years. [72,73]

The approach described above to estimating mortality attributable to influenza in a particular influenza season is limited in two ways. First, the epidemiologist must decide what observations in the previous five years of P&I data are to be deleted before model fitting. The uncertainty in

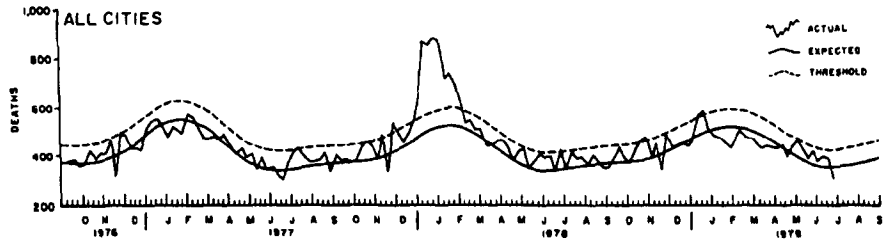


FIGURE 1. Observed number of weekly deaths attributed to pneumonia and influenza, compared with number forecast by Serfling regression model (1), 121 U.S. cities, 1976–1979.

Figure 2.6.: **The 'epidemic threshold' in P&I data from 121 US cities between 1976/77 and 1978/79 calculated using Serfling's least squares method (figure 1 from [74]).**

the designation of observed mortality or morbidity as 'epidemic' or 'non-epidemic' (i.e. 'aberrant' or 'normal') is not taken into account. Second, the roughly sinusoidal baseline with a linear (or sometimes quadratic) long-term trend lacks flexibility in fitting 'normal' seasonal incidence when the timing of influenza seasons varies between influenza seasons.

As an example of a recent adaptation of Serfling's original method, Simonsen *et al.* used a Serfling-type regression model to estimate excess P&I and all-cause (AC) mortality in the elderly in the US for influenza seasons between 1968 and 2001. [7] First, rates of P&I and AC were adjusted for the change from ICD 9 to ICD 10 using a comparability ratio. Any long-term trend in the mortality rates was removed by dividing each month's rate by the average summer (June - August) mortality rate using a smoothing spline. Mortality rates for December to April where discarded to isolate 'normal' mortality. A Serfling-type regression model was fitted to de-trended monthly rates of influenza-registered mortality, excluding December to April, according to the formula

$$D_t = \alpha + \beta_1 \sin\frac{2\pi t}{12} + \beta_2 \cos\frac{2\pi t}{12} + \epsilon_t$$

where D_t is the de-trended monthly mortality rate, α is the intercept, β_1 and β_2 are coefficients for seasonality and ϵ_t is the error term. Months when the observed rates of influenza-registered death exceeded the upper 95% CI on the model-predicted rates of influenza-registered death were defined as

'epidemic months'. The identical Serfling-type regression model was then fitted to de-trended P&I and AC in 5-year age groups (65-94). Excess P&I or AC was the difference between observed and model-predicted P&I or AC for all 'epidemic months' with trend added back in. This model improves upon Serfling's original least squares approach by more flexibly accounting for long-term trend in 'normal' mortality data using a smoothing spline. The models described in the thesis fitted seasonality using a Fourier term (one sine plus one cosine term) and fitted long-term trend with cubic splines. This is analogous to the method of Simonsen *et al.* where seasonality was modeled using a Fourier term and where long-term trend was removed using a smoothing spline. The models described in the thesis thus have increased flexibility for fitting seasonality and trend in the data relative to Serfling-like least squares.

The adaptation by Simonsen *et al.* of the Serfling method is still limited by the need to delete observations for December to April before model fitting. Accordingly, the models described in chapters 5 to 7 of the thesis have differentiated 'aberrant' from 'normal' morbidity and mortality as part of model fitting, while incorporating uncertainty into this differentiation.

2.4.2. ARIMA time series

Choi and Thacker proposed a method to increase the accuracy of forecasts, and estimates of excess mortality, compared to those from the Serfling-like least squares method. [74] They adapted a seasonal autoregressive integrated moving average (ARIMA) time series model for this purpose. [74, 75] Like Serfling's method, the seasonal ARIMA model was fitted to past P&I data with 'aberrant' (or 'epidemic') weeks deleted. Unlike Serfling's method, in the ARIMA model non-independence in the P&I data is not assumed to necessarily follow a roughly sinusoidal pattern plus linear trend. The ARIMA model predicts expected numbers of deaths for future time periods based on the temporal sequence of counts of deaths in past time periods. Baseline counts for deleted epidemic weeks are also predicted as part of model fitting. Comparing figure 2.7 to figure 2.6, the increased flexibility of the seasonal ARIMA model compared with Serfling least squares model for fitting 'normal' incidence is evident. Excess mortality was defined as the sum of positive residuals exceeding 1.65 SD above mortality predicted by

the ARIMA model.

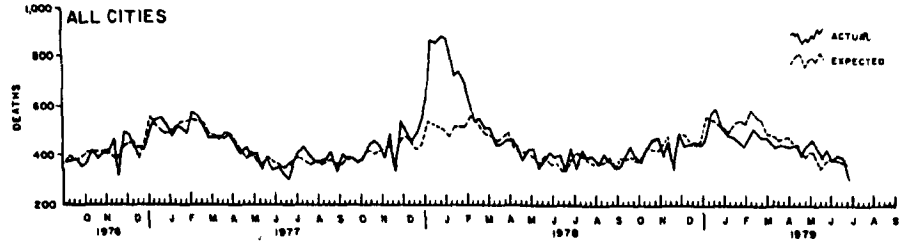


FIGURE 2. Observed number of weekly deaths attributed to pneumonia and influenza, compared with number forecast by ARIMA model (2), 121 U.S. cities, 1976–1979.

Figure 2.7.: **The 'epidemic threshold' in P&I data from 121 US cities between 1976/77 and 1978/79 calculated using the ARIMA time series method (figure 2 from [74]).**

The seasonal ARIMA time series model for P&I developed by Choi and Thacker produced more accurate estimates of excess P&I than Serfling-like least squares [74] because previous 'aberrant' observations were replaced by model-predicted counts instead of simply being deleted. The models described in chapters 5 to 7 of the thesis did not require deletion of data prior to model fitting. The ARIMA method is limited in a similar way to the Serfling method in requiring the deletion of 'aberrant' observations before model fitting. Both the ARIMA and Serfling-like least squares methods assume that excess mortality is entirely attributable to influenza.

2.4.3. Rate difference

Another general approach to differentiating 'normal' from 'aberrant' incidence is to calculate rates of morbidity and mortality during the 'aberrant' period (defined based on the presence of an influenza virus indicator) and compare these to rates during a 'peri-seasonal period' (winter weeks outside of those defined as 'aberrant') or during the summer. [76] As an example of this 'rate difference' method, Jansen *et al.* recently compared average rates of excess all-cause mortality in elderly people in the Netherlands in periods of high and low vaccine coverage in order to investigate vaccine impact on mortality. [64] Excess all-cause mortality was estimated using the rate difference method. First 'influenza active periods' and 'RSV active periods' in each influenza season under study were identified. 'Influenza active

weeks' were defined as $\geq$ 2 consecutive weeks accounting for at least 5% of the winter's total influenza laboratory reports. 'RSV active weeks' were $\geq$ 2 consecutive weeks accounting for at least 5% of the winter's total RSV laboratory reports. Weeks of influenza predominance were 'influenza active weeks' which were not 'RSV active weeks'. Winter was considered to last from week 40 of one calendar year to week 20 of the following year. The 'peri-seasonal baseline' mortality was AC mortality in winter weeks which were neither 'influenza active weeks' nor 'RSV active weeks'; the 'summer baseline' mortality was AC mortality in weeks 21 to 39. Excess AC mortality rates for each season were calculated by subtracting rates during the peri-seasonal baseline from rates during periods of influenza predominance, multiplied by the number of 'influenza active weeks' in that season. Excess was similarly calculated using the summer instead of peri-seasonal baseline. Estimates of excess mortality using this method may be biased downwards if influenza - attributable deaths lag laboratory reports by a number of weeks (thus falling outside of the virus active periods).

Many of the estimates of excess morbidity and mortality in England & Wales summarised in section 2.3.1 were calculated using a similar method to that of Jansen *et al.*. [1,2,6,49,50,53,57] In these studies, an 'epidemic threshold' was defined using average all-age ILI data for winter weeks when there were no laboratory reports for influenza (routine or sentinel). Winter weeks when observed ILI was above the 'epidemic threshold' were defined as 'influenza active weeks'. Average observed morbidity or mortality in 'influenza active weeks' was divided by average morbidity or mortality for the winter weeks of that season which were not 'influenza active weeks', multiplied by the number of 'influenza active weeks', to give total excess morbidity or mortality in that season. The logic behind multiplying a ratio of average rates by the length of the influenza active period is unclear. The 'influenza active weeks' (defined using all-age ILI data and all-age influenza A and B laboratory reports) were assumed to be the same for all age groups and for all outcome variables (e.g. ILI, respiratory hospitalisations and mortality). This assumption would not hold if influenza activity increased in some age groups before others, or if deaths lagged ILI (see section 3.2.6). Also, for analysis of a number of influenza seasons at once, estimating the 'epidemic threshold' by averaging ILI rates across weeks for several seasons would mean the 'epidemic threshold' was too high or too low at the ends of

the time series if there were a long-term increase or decrease in ILI incidence (see section 4.6).

2.4.4. Regression on an indicator of influenza virus circulation

Another approach to estimating mortality (or morbidity) attributable to influenza is to regress non-laboratory confirmed outcome incidence on indicators of influenza virus circulation (like laboratory reports for influenza A and B, GP consultations for ILI or deaths registered to influenza). By doing this, incidence attributable to influenza is distinguished from incidence attributable to other factors, like seasonality, fluctuating ambient winter temperatures or epidemics of RSV. Clifford *et al.* introduced this approach by fitting multiple linear regression models to 4-weekly deaths (all-cause or respiratory in separate model fits) according to the formula

$$Y_t = \alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \epsilon_t$$

where Y_t is the 4-weekly number of deaths, α is the intercept and βs are coefficients for dummy variables for season, trend, ambient temperature, years since an antigenic drift event, RSV and various indicators of influenza virus circulation that included influenza A and B routine laboratory reports and GP consultation rates for influenza from the RCGP. [52] Coefficients were estimated using least squares. Clifford *et al.* estimated the number of deaths in the absence of influenza by setting the influenza indicator to zero (or the lowest observed level in the case of the RCGP consultation rate for influenza, which is rarely zero). The portion of excess mortality attributable to influenza was then the observed mortality minus that predicted when the influenza indicator was set to the appropriate baseline level.

This general approach has been widely used: Poisson models with a log-link [8, 77] or an identity-link [78] have also been fitted. These regression models can also incorporate terms for observed morbidity or mortality in previous weeks or months to account for dependence between death counts not accounted for by variability in the indicator of influenza virus circulation, seasonality, trend and other confounders. [42] Carrat and Valleron merged the regression and ARIMA approaches when they regressed respi-

ratory, cardiovascular and other mortality rates on rates of mortality registered to influenza and an error term that had an ARIMA structure. [79] In this way only the variability in respiratory, circulatory or other mortality rates that could be explained by variability in rates of mortality registered to influenza was attributed to influenza. All of the remaining variability in the dependent variable was accounted for in the ARIMA process. In the general class of models described in this section, influenza - attributable mortality or morbidity is estimated either as observed mortality minus mortality when the influenza indicator is set to its baseline value [52] or multiplying the regression coefficient for the influenza virus indicator by the magnitude of the influenza virus indicator in each week of each influenza season. [9] There is a conceptual issue and an analytical limitation of regressing surveillance data on each other.

The conceptual issue with regressing different influenza surveillance data on each other is that true influenza burden is partially observed in each of the surveillance data sets (e.g. mortality, GP consultations, laboratory reports for influenza) (figure 2.8). When surveillance data are regressed on one another, this implies that the association between the data sets is as in figure 2.9. That is, the implication is that laboratory-confirmed cases (e.g. ascertained through laboratory reports for influenza) cause non-laboratory confirmed cases (e.g. deaths coded to underlying pneumonia or influenza). As mentioned above, in reality, influenza disease is partially observed in each data set. This suggests multivariate models, where both non-laboratory-confirmed and laboratory-confirmed data are outcome variables, as an alternative to regression models.

The analytical limitation of regressing different streams of influenza surveillance data on each other is that convenient aspects of models with a log-link (e.g. allowing control for the changing size of the population at risk with a population offset, see chapter 3) are problematic. Poisson models with a log-link and a population offset imply a multiplicative association between the influenza virus indicator (e.g. laboratory reports for influenza) and the dependent variable (e.g. deaths coded to underlying pneumonia or influenza). This may not be realistic. [80] Other scales of analysis (for example, modeling the association between the laboratory reports for influenza and deaths coded to underlying pneumonia or influenza using a Poisson model with an identity-link or using a Gaussian model) make adjustment for the changing

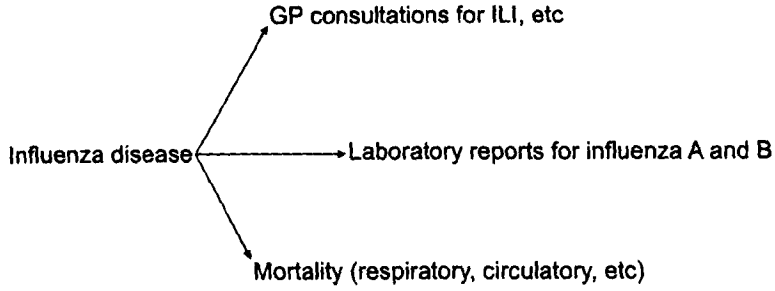


Figure 2.8.: **A schematic illustrating the implied relationship between true influenza incidence and influenza morbidity and mortality as captured by surveillance data.**

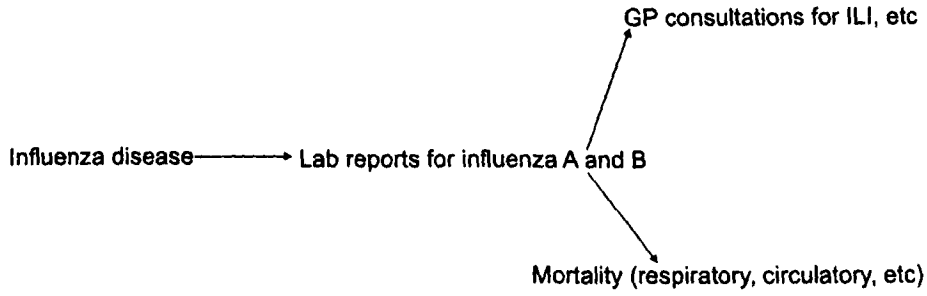


Figure 2.9.: **A schematic illustrating the implied relationship between true influenza incidence and influenza morbidity and mortality as captured by surveillance data when GP or mortality data are regressed on laboratory report data.**

size of the population at risk more difficult because a population offset cannot be used. In Gaussian models, counterintuitive negative model-predicted numbers of cases are possible (see chapter 3).

Thompson *et al.* recently estimated excess respiratory and circulatory morbidity in the US between 1977/78 and 2002/03 using four methods described above: Serfling-like least squares, ARIMA time series, rate difference and Poisson regression with a log-link (where seasonality was modeled using a sine and cosine term). [60] The highest estimates of influenza - attributable mortality were derived from the rate difference method using a summer baseline and the next highest from the rate difference method using a peri-seasonal baseline; Serfling least squares, Poisson regression with a log-link and ARIMA time series models gave similar estimates to each other and lower estimates than the rate difference method.

2.4.5. Latent variable time series models

An alternative way to distinguish influenza - attributable from non-influenza attributable morbidity and mortality is to use latent variable time series models (details in section 3.5.2). In latent variable time series models, the distinction of influenza - attributable ('aberrant') from non-influenza - attributable ('normal') morbidity and mortality is automated. The model is able to distinguish 'aberrant' from 'normal' morbidity and mortality by considering the morbidity or mortality time series as having arisen through a latent (unobserved) process (e.g. [81]). The differentiation of 'aberrant' from 'normal' incidence is governed by latent 'states' where each state has a probability distribution associated with it. The model learns from the data about the probability that any given week is drawn from the 'normal' or 'aberrant' distribution. The contribution of the work described in the thesis is in developing age group-specific multivariate latent variable time series models for influenza morbidity and mortality. In the models described in chapter 6, the distinction of 'aberrant' from 'normal' P&I and ILI is informed by laboratory reports for influenza A. Age group specific laboratory reports, P&I and ILI data are simultaneously modeled as outcome variables. The distinction of 'aberrant' from 'normal' incidence in P&I and ILI is informed by both laboratory reports for influenza A and the P&I and ILI data themselves. In this way the specificity of influenza - attributable

morbidity and mortality for influenza is increased relative to models fitted to P&I or ILI only (see chapter 5). Because models are age group-specific, different timing of influenza seasons in different age groups is allowed for. Also, because in bivariate models the distinction of 'aberrant' from 'normal' incidence is informed by P&I, or ILI, and laboratory reports, shifted timing of 'aberrant' incidence in ILI compared to P&I data is accommodated.

In the latent variable framework, the observation model for the 'normal' incidence probability distribution has often been modeled as Serfling-like with Poisson or Gaussian errors. [81–83] Using a Gaussian model is limited because of the possibility of predicting negative counts. As mentioned previously, throughout the thesis 'normal' incidence was fitted with a Serfling-like Poisson model.

In previous latent variable time series models, the distinction between 'normal' and 'aberrant' incidence has been captured in an autoregressive term [83–85] or as an additive mean shift. [82] Rath *et al.* chose an exponential distribution for the 'normal' incidence observation model and a Gaussian distribution for the 'aberrant' incidence observation model. [86] An advantage of this approach was that seasonality and trend did not need to be explicitly modeled. There is not an obvious choice as to how to model the difference between 'normal' and 'aberrant' incidence. The models described in chapters 5 to 7 modeled this using a random effect mean shift for each influenza season (see section 3.5.2). The random effect allows variability in the intensity of influenza seasons. The random effect mean shift is a simple first step for modeling the relative impact of influenza seasons in terms of morbidity and mortality.

2.4.6. Transmission dynamic models

An alternative to the statistical models described above for making inference about a time series of counts of an infectious disease are transmission dynamic models. An example of this approach is the model introduced by Finkenstadt *et al.*: a susceptible-infected-recovered-susceptible (SIRS) model where an observed time series of counts of the disease in question populates the 'infected' class. [87] This framework was recently applied to French ILI data. [88] The weekly ILI count was assumed to arise from a Gamma distribution whose mean (ϕ_t) was expressed in terms of the contact

rate (η) and the number susceptible (ζ) and infected (κ) in week t-1. [88]

$$\phi_t = \eta_t \zeta_{t-1}^{\chi} \kappa_{t-1}$$

χ is a mixing parameter for the contact process between infected and susceptible people. χ takes the value 1 if the population mixes homogeneously. Seasonality was parameterised by allowing η to vary seasonally (with a period of 1 year). The number susceptible (ζ) and recovered (ω) in week t+1 was then

$$\zeta_{t+1} = \zeta_t - \kappa_{t+1} + \gamma_t \omega_t$$

$$\omega_{t+1} = (1 - \gamma_t)\omega_t + \kappa_{t-1}$$

The parameter γ_t captured the return of immune individuals to susceptibility because of waning immunity or due to antigenic drift. In the full version of this model, γ was constant throughout an influenza season except for at most 1 week when it could take any value up to 1. The 1 week during which a proportion of the population essentially instantaneously reverted from immune to susceptible represented the introduction of an antigenically drifted variant. Inference about unknown quantities (e.g. χ, η, γ) was made using Markov chain Monte Carlo (see section 3.5.3).

Latent-variable models can be extended to include a transmission dynamic component [89](where the Markov chain, of which the observed data are a realisation, is generated by a transmission model). In the work described in the thesis, a strictly statistical, and not a transmission dynamical, framework has been used.

2.5. Determinants of variable impact of influenza seasons

In this section, key potential determinants of the variability in influenza - attributable morbidity and mortality between influenza seasons are introduced. These potential determinants include the exposures of interest in the work described in the thesis: large antigenic drift events in influenza A/H3N2 virus evolution and vaccination.

There are a number of factors that help to explain the variability in influenza - attributable morbidity and mortality between influenza seasons.

One is the dominant circulating variant or variants in each season. Influenza A/H3N2 virus infection generally causes more serious illness than infection with influenza A/H1N1 or B virus. [40,90,91] Influenza seasons dominated by influenza A/H3N2 virus result in higher mortality [1,92], hospitalisations [1,48] and GP consultations for influenza-like-illness [1] than seasons dominated by influenza A/H1N1 or B viruses. There is evidence that the H3 haemagglutinin is under positive selection, [11,12] that there is a higher mutation rate of influenza A/H3N2 virus vs. H1N1 or B viruses [12] and some evidence of more efficient transmission of influenza A/H3N2 than H1N1 viruses. [19] Large antigenic drift events in influenza A/H3N2 virus evolution often coincide with epidemics (see section 2.6). [93,94]

Increasing vaccine coverage of certain age groups may lead to less influenza - attributable morbidity and mortality through direct and indirect vaccine effects (see section 2.8). [95] The mismatch between vaccine and dominant viruses sometimes results in lower VE and should be taken into account in studies of the association between antigenic drift, vaccine coverage and excess morbidity and mortality. [16,17]

Factors not related to influenza directly, like ambient temperature and RSV epidemics, [42] a decline, and then leveling off, of levels of smoking in the population [96] and declines in healthcare-seeking behaviour in the population [97] may explain some of the variability in influenza - attributable morbidity and mortality influenza season to influenza season. Factors such as these may obscure the relationship between influenza-related factors (e.g. antigenic drift or vaccine coverage) and excess morbidity and mortality.

2.6. Plausibility of an effect of antigenic drift on excess morbidity and mortality

In this section, the plausibility of an average inflating effect of large antigenic drift events on mean influenza - attributable morbidity and mortality is discussed.

2.6.1. Evolution of influenza A/H3N2 virus

Laboratory studies have provided evidence for changes in influenza A/H3N2 virus that may be indications of adaptation of influenza A/H3N2 virus to

transmission within the human population. That is, these changes may be due to selective pressures on influenza A/H3N2 virus exerted during the long-term circulation of influenza A/H3N2 virus in the human population (41 years). For example influenza A/H3N2 viruses have, over time, become better able to bind sialic acid receptors in human epithelial cells. [98] They also appear to have developed a lower affinity for natural killer cells. [99] Because this evidence is not coupled to information about viral fitness, however, it is difficult to interpret it in terms of resulting changes to virulence (the severity of the illness caused by infection with a pathogen). For example, mutations that increase sialic acid affinity, and enhance host cell binding, might also increase antibody recognition and result in an overall decrease in fitness. [98]

Antigenic evolution

Study of the antigenic evolution of influenza A/H3N2 virus gives an indication of evolution in response to selective pressures by human immune systems which does not result in loss of fitness. Less fit viruses do not persist in order to be isolated from large numbers of patients and thus are not well represented in databases of antigenicity data on influenza viruses.

As part of the influenza vaccine strain selection process, the World Health Organization (WHO) assesses antigenicity using the haemagglutination inhibition (HI) assay. The HI assay is based on the ability of influenza viruses to agglutinate red blood cells and the ability of specific antibodies to inhibit this agglutination. The antibodies used in the HI assay are raised by infecting ferrets with the relevant influenza virus variants. The HI value, or titre, ascribed to each antiserum is the highest dilution of the antiserum that can block the agglutination of the red blood cells. The higher the HI value, the more similar the variant of interest is to the reference variant against which the antiserum was raised. The HI test is used for surveillance because the antibody produced is durable and directed towards the haemagglutinin which is the most relevant for assessing susceptibility and immunity. [100] These data have historically been used in a qualitative way by the WHO to estimate how well a vaccine that included one variant should protect vaccinees from infection with another variant.

Recently, Smith *et al.* used HI assay data on 79 antisera and 273 viral

isolates from the WHO vaccine strain selection data set to identify clusters of antigenically similar variants. [13] An antigenic map was constructed by placing virus isolates and antisera from influenza seasons between 1968 and 2003 on an x-y plane. The distance between virus-antiserum pairs was determined by their HI titre. The relative position of each pair on the plane was determined by minimising the sum of squared differences between the $\log_2$ of the HI value and the physical distance between the antigens and antisera on the map (figure 2.10). The accuracy of the map was determined by selecting antigen-antiserum pairs whose HI values were not included in the creation of the map, predicting the HI value of those pairs from distances in the map, and then testing the HI value for this pair in the laboratory. The correlation between HI distance inferred from the map and laboratory-tested distance was 0.81.

The HA1 subunit of the HA of each of the influenza A/H3N2 virus variants included in the map was sequenced to allow comparison of the rates of genetic and antigenic change. Smith *et al.* provided evidence that viral genetic evolution happens in a relatively continuous (linear) fashion while antigenic change is punctuated. [13] This was done by plotting both genetic distance (in amino acid substitutions) and the antigenic distance between the dominant variants each season against the season (figure 2.11). A cluster of variants in the antigenic map was dominant for a mean of 3.3 years (range 1 to 8 years) before a large antigenic drift event, or cluster transition (CT), occurred. There was some overlap in circulation of adjacent clusters.

More recently, Russell *et al.* showed that antigenic change in influenza A/H3N2 virus between 2002 and 2007 was more linear [101] than the historical average indicated by Smith *et al.*. [13] The analysis of antigenic evolution between 2002 and 2007 was based on a larger data set that was more geographically representative of global isolates; the earlier analysis of antigenic evolution between 1968 and 2003 was based on a data set where 94 of 273 isolates used to generate the antigenic map were from the Netherlands. An alternative hypothesis to the rate of antigenic change having been more linear between 2002 and 2007 than previously is that global antigenic change is gradual compared with local antigenic evolution. Local evolution may appear punctuated because influenza virus does not persist locally between influenza seasons. [101]

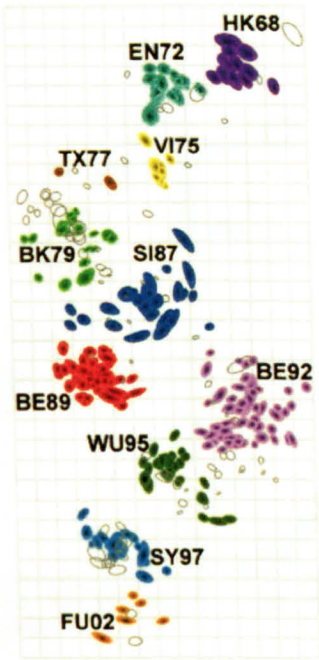


Figure 2.10.: **Antigenic map showing clustering of antigenic variants of influenza A/H3N2 virus between 1968 and 2003/04. Reproduced from figure 1 of Smith *et al.* [13].** One unit of antigenic distance on the map is equivalent to a 2-fold dilution in the antiserum in the HI assay. Coloured symbols are viral isolates and open symbols are antisera. The size and shape of each point on the map reflects a confidence area in its placement on the map. Antigenic clusters are distinguished by colour. Note that clusters of antigenically similar viruses arrange chronologically from the HK68 cluster, that includes the variant that caused the pandemic of 1968-1970 (top), to the FU02 cluster.

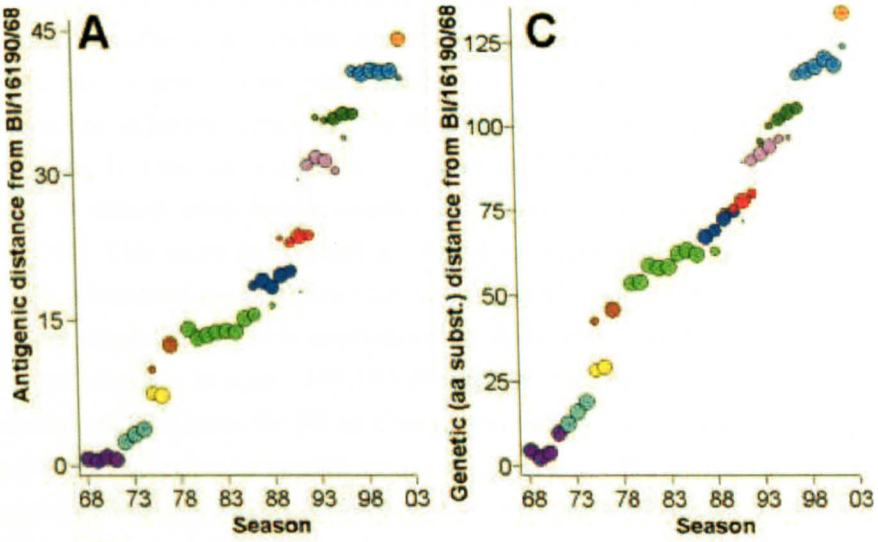


Figure 2.11.: **Antigenic distance (left) and number of amino acids (right) between circulating influenza A/H3N2 virus variants each season against the season (reproduced from figure 4 in [13]).** Clusters are colour-coded as in the previous figure. The area of each point represents the proportion of isolates per season which were from each antigenic cluster. Sampling is biased towards outliers so the area of each point does not reflect this variant's epidemic impact. Antigenic change appears to proceed in a more punctuated manner than genetic change.

2.6.2. Relationship of antigenic drift to host immunity

There is evidence that the HA1 subunit of influenza A/H3N2 virus is under positive selection and it is likely that a major driver of this is host population immunity. [11,12] Using a transmission dynamic model, Gog *et al.* showed that strain-specific immunity can lead to CT seasons. [102] Models that also allow a non-specific strain-transcending immunity reproduce antigenic evolution that proceeds gradually but also clustered antigenic evolution like that observed by Smith *et al.*. [12,13]

Antigenic drift plausibly increases the size of the susceptible population to circulating variants of influenza virus. This is because antibodies generated by natural infection or vaccination against previously circulating variants become less able to neutralise variants which are antigenically drifted. As a result, people revert from being immune to being susceptible to circulating variants of influenza virus. A cohort study where individuals were grouped according to previous exposure to influenza A/H3N2 viruses suggests that there is almost complete immunity to variants within an antigenic cluster. [103] This same study, and a natural experiment in military personnel where vaccinated people were housed with newly arrived unvaccinated individuals, suggested there is approximately 60 to 80% cross immunity between clusters adjacent in time. [103,104] Nakajima *et al.* found no cross-immunity between viruses from the HK68 cluster and serum from human subjects that included antibodies to viruses circulating between 1991 and 1993. [105] This suggests there probably is little or no cross-immunity between non-adjacent clusters.

The make-up of the trivalent influenza vaccine used in most countries is updated regularly to track the antigenic evolution of (specifically) influenza A viruses. [106] The influenza A/H3N2 virus vaccine variant is updated when there is an antigenic distance of at least 2 units (a fourfold dilution of antiserum in the HI assay) between the vaccine variant and the variant expected to circulate in the next influenza season. The mean degree of antigenic change with each cluster transition is 4.5 antigenic units. [13] There is at least 1 vaccine variant in each influenza A/H3N2 virus antigenic cluster (see section 2.6.1).

2.7. Evidence for effect of antigenic drift

In this section, the evidence for an average inflating effect of large antigenic drift events on influenza - attributable morbidity and mortality is reviewed.

The appearance of antigenically distinct variants of influenza A/H3N2 virus in 1972 and 1975 were accompanied by large increases in deaths from influenza and pneumonia in the United Kingdom (UK) and worldwide. [93] Greene *et al.* plotted the monthly percentage AC that was due to P&I from 1968 to 1998 for US residents aged 65 and over, an indicator of the relative impact of influenza seasons in terms of mortality, indicating CT seasons on the graph (figure 2.12). [94] I added the red horizontal line to their plot to indicate approximately the average epidemic in terms of percentage P&I. All CT seasons coincided with an average or above average P&I percentage though not all of the highest peaks in the graph occurred during CT years. The percentage of all-cause deaths coded to pneumonia or influenza is sensitive to changes in the relative percentages of other causes of death. As such, it has not been used in work described in the thesis as an indicator of the relative impact of influenza seasons.

Most reports in the literature which make mention of the impact of large antigenic drift events in influenza A/H3N2 virus evolution refer to individual drift events [63, 107, 108] or to a number of antigenic drift events that coincided with massive epidemics. [58,93,109,110] The average effect of cluster transitions on morbidity or mortality has not, to my knowledge, been quantified. There may therefore be a bias in the published literature towards large antigenic drift events in influenza A/H3N2 virus which resulted in large excess mortality and/or morbidity.

The best estimate of the average effect of a number of CT seasons comes from a model of the antigenic evolution of influenza A/H3N2 virus coupled with a transmission dynamic model. [14] Here antigenically similar variants were modeled as belonging to neutral ensembles and the individuals as being susceptible to, infected by or recovered from infection from a neutral ensemble instead of a particular variant. This model reproduced the pronounced seasonality in incidence and duration of cluster dominance observed in reality. Reading peak attack rates for each season from a figure of predicted incidence, colour-coded for the dominant influenza A/H3N2 virus cluster (figure 2.13), model-predicted influenza incidence during the first season

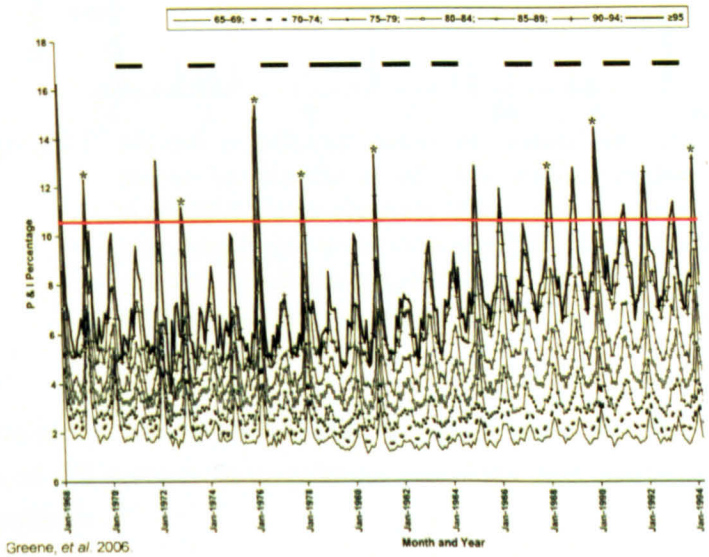


Figure 2.12.: **The percentage of AC in those 65 and over in the US coded to underlying P&I, influenza seasons 1968/69 to 1997/98. Stars indicate seasons which were the first influenza A/H3N2 virus-dominated season after a CT.** Adapted from [94]. The black horizontal bars are seasons dominated by influenza A/H1N1 and/or influenza B viruses. The red horizontal line indicates the approximate mean peak height in terms of P&I percentage (not in the original paper). All CT seasons had an average or above average P&I percentage.

after a new cluster emerged was approximately 1.6 times higher than the average incidence in other seasons.

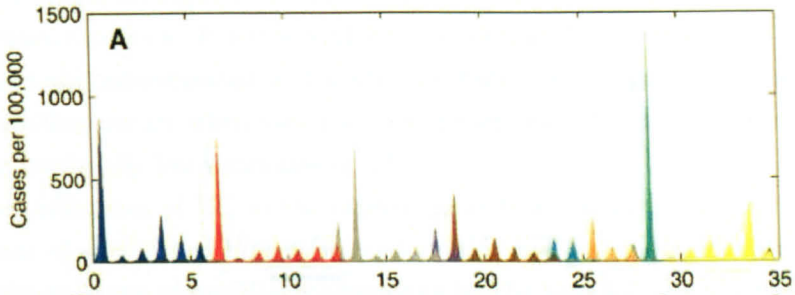


Figure 2.13: **Model predicted rates of 'cases' of influenza from a paper by Koelle *et al.* [14], colour-coded by influenza A/H3N2 virus cluster.** The y-axis is case rate and x-axis is year, numbered from 1968 to 2003. Model-predicted antigenic clusters of influenza A/H3N2 virus are dominant for 1 to 6 influenza seasons (vs. 1-8 influenza seasons for clusters observed in reality).

Except in the descriptive and theoretical ways described above, the effect of CT seasons on population morbidity and mortality has not been quantified.

2.8. Plausibility of impact of vaccination of the elderly

In this section, the plausibility of an impact of increasing yearly vaccine coverage of the ≥ 65 age group on mean influenza - attributable morbidity and mortality is discussed.

2.8.1. Vaccine efficacy in the elderly

Trivalent inactivated influenza vaccine has been shown to prevent up to 58% (95% CI 26-77%) of laboratory confirmed clinically diagnosed ILI in healthy people aged ≥ 60. [15] Vaccine efficacy of trivalent inactivated vaccine against confirmed influenza in healthy younger adults is approximately 77% (95% 67-85%). [111] A recent review suggested similar post-vaccination antibody levels in the elderly as in younger adults. [112] This is in contrast

to the general thinking that antibody response to vaccination declines with age.

2.8.2. Vaccine effectiveness in the elderly

Estimates of VE in the elderly are complicated because few studies used laboratory-confirmed outcomes. There is the potential for both positive and negative confounding in observational studies that have not used laboratory-confirmed outcomes. Positive confounding occurs when vaccinees are healthier than the unvaccinated and leads to inflated estimates of VE. Negative confounding occurs when vaccinees are frailer than the unvaccinated, leading to artificially low estimates of VE.

Best estimates of VE in the elderly come from studies that verified the absence of positive and negative confounding by health status. For example, the presence of positive confounding by the healthy vaccinee effect can be revealed by looking for VE in the pre- or post-influenza season (either before influenza vaccine has been administered or after the end of influenza virus circulation, when VE would be expected to be zero). From studies not affected by positive or negative confounding, VE against acute respiratory hospitalisations is approximately 20-30% [113, 114] and against respiratory mortality is between 12% (95% CI 8-16%) [113] and 79% (95% CI 0-100%). [77]

2.8.3. Indirect effect of vaccination

Several studies have shown that influenza vaccination of school children (e.g. [115]), health care workers (e.g. [116]), or whole towns (e.g. [117]) can provide indirect protection against ILI and mortality in the unvaccinated.

In England & Wales, influenza vaccination is recommended for people over 65 years of age and for people <65 with risk factors for complications of influenza (e.g. respiratory, circulatory diseases and diabetes) [118]; the majority of people receiving the vaccine are over the age of 50. [119] If vaccine efficacy against laboratory-confirmed clinically diagnosed influenza is approximately 58%, vaccine effectiveness against non-laboratory confirmed ILI should be lower. This is because ILI caused by pathogens other than influenza would be expected to be equally distributed between the vaccinated and unvaccinated elderly, biasing estimated VE downwards. (See work de-

scribed in chapters 5 and 6 to increase the specificity of non-laboratory confirmed P&I and ILI for influenza by jointly modeling P&I and ILI with laboratory reports for influenza A). Yearly vaccine coverage of the $\geq$ 65 in England & Wales increased from 24% in 1989/90 to 71% in 2004/05 (and from 11% to 40% in people <65 in defined risk groups). [119,120] Unless coverage is high enough to elicit substantial herd immunity, it will be difficult to demonstrate an impact of the vaccine on levels of non-specific population markers of influenza disease (see chapter 7). In addition, drivers of transmission in households and the community are thought to be children. [38] Observing a herd effect of vaccinating the elderly on transmission in other age groups is unlikely because of the probably low transmission from elderly to other age groups.

2.9. Evidence for the impact of vaccination of elderly

In this section, the evidence for an impact of increasing yearly vaccine coverage of the $\geq$ 65 age group on mean influenza - attributable morbidity and mortality is reviewed.

2.9.1. Descriptive studies

In several settings there has been a crude decline in rates of respiratory mortality, GP consultations or emergency department visits in the elderly in the community, concurrent with an increase in vaccine coverage of the elderly (detail in table 2.2). There may be an impact of increasing vaccination of the elderly on rates of respiratory mortality, GP consultations or emergency department visits but these reports do not provide strong evidence to support this. This is because estimates of trends in total (not excess or influenza - attributable) morbidity or mortality are sensitive to changes in baseline rates of morbidity and mortality. For example, Dijkstra *et al.* regressed total ILI rates from influenza seasons on vaccine coverage of the $\geq$ 65 in those influenza seasons and provided some evidence of negative correlation between the two. [121] A decline in use of GP services, like that observed in recent years in England & Wales [122], would give the impression of a causal relationship between increasing vaccine coverage and

declining rates of ILI when part or all of the decline in ILI rates may be due simply to declining use of GP services. Because of the potential confounding by changes to baseline morbidity or mortality incidence, as well as possible confounding by antigenic drift, the dominant circulating variant, or effect modification by the dominant variant or vaccine mismatch, these descriptive studies cannot be interpreted as showing strong evidence of vaccine impact.

Table 2.2.: **Descriptive studies of the impact of changing vaccine coverage of the elderly on morbidity and mortality.** In the table, season(al) refers to an influenza season.

Setting	Years	Exposure	Methods	Findings	ref
Alberta, Canada	1999/00-00/01	coverage (range 27-77%) in 17 districts, by season	Pearson's correlation coverage vs. pneumonia GP, hosp., mortality rates	-0.59 1999/00, -0.79 2000/01 (GP and hosp.), smaller correlation for mort.	[123]
US	1992/93-01/02	coverage (range 29-78%)	pneumonia hosp. rates by fiscal year (Oct-Sept)	10.7/10,000 (1992) to 5.2/10,000 (2001)	[124]
US	1992/93-97/98	coverage (50-67%)	plot P&I rates by season	stagnant 1994-1996, down 1997	[125]
Stamford, US	1996/97-01/02	# vaccinations (range 7,387-18,471), coverage unclear	Chi-square test for trend all-age resp. or chronic obstructive pulmonary disease (COPD) A&E visits, resp hosp., P&I	resp. A&E 144.36 ($p<0.001$) , COPD A&E 31.15, $p<0.001$ (little change resp hosp., no change P&I)	[126]
Wisconsin, US	1980-06	coverage (range 0-72%)	age-group-specific P&I rates (0-4, 5-49, 50-64, 65-74, 75-79, 80-84, $\geq$ 85) by ?calendar yr.	all age groups lower rate 2001-2003 than 1980-1982, except $\geq$ 85 (27% rise)	[127]
Holland	1991/92-04/05	coverage (range 30-$\geq$70%)	Spearman's rank correlation, linear regression seasonal ILI (?all-age or $\geq$ 65)	-0.62 ($p=0.01$); -1.7/10,000 per % change in coverage (95% CI -3.3 to 0.01)	[121]
Sao Paulo state, Brazil	1980-00	elderly vaccination from 1999, coverage unclear	extrapolated (to 1999,2000) linear regression age group-specific resp. mort for 1980-1998	observed mort in 2000 <lower bound predicted women (all ages), men (70-74, $\geq$80)	[128]
Portugal	1998/99-06/07	coverage (range 31-50%)	correlation total winter ILI rates with coverage	all seasons: Pearson $r=-0.534$, $p=0.173$; Spearman $rs=-0.359$, $p=0.382$; H3N2-dominated seasons: Pearson $r=-0.911$, $p=0.011$; Spearman $rs=-0.899$, $p=0.015$	[129]
Taiwan	1997-02	coverage (range 2-62%)	plot annual death rates (pneumonia, other resp. causes); coverage overlaid	decline resp. death rate (no clear trend pneumonia death rates)	[130]

2.9.2. Regression of excess mortality on vaccine coverage

Seven studies were identified which assessed the association between excess or influenza - attributable respiratory hospitalisations or mortality in the elderly on vaccine coverage in the elderly (detail in table 2.3). Excess influenza - attributable morbidity or mortality is less sensitive to changes in baseline morbidity or mortality than total morbidity or mortality rates. As such, these studies provide stronger evidence for or against an impact of vaccination of the elderly than the descriptive studies in table 2.2.

Table 2.3.: **Studies which assessed changing excess morbidity or mortality with changing vaccine coverage of the elderly.** In the table, season(al) refers to an influenza season.

Setting	Years	Coverage	Methods	Adjustment	Findings	Notes	ref
US	1968/69-00/01	16% 1972, 65% 1999-01	Trend in $\log_{10}$ seasonal P&I/AC	Age, age*year, by H3N2 seasons	-12% P&I 2001 vs 1980, lower bound 95% CI -35%	residuals autocorrelated; overdispersion; confounding by temp., antigenic drift?	[7]
Italy	1970/71-00/01	low (<10% pre-1986), rising (10-61% 1987-01)	Trend in $\log_{10}$ seasonal P&I/AC	Age, age*year, by H3N2 seasons	-9% P&I per year 1970-86; +2% P&I per year 1987-01 (similar AC)	no CIs; confounding by temp., antigenic drift?	[67]
Sao Paulo, Brazil	1993-02	0 pre-1998, 67% 2000	Average P&I rate (Serfling or ARIMA), 1998-02 vs. 1993-97	Age, sex	-47% (Serfling), -88% (ARIMA) P&I 1998-02 vs. 1993-97	no CIs; confounding by H3N2, antigenic drift?	[131]
Finland	1992/93-93/94	20% area 1, 48% area 2	Hosp. (pneumonia, influenza and circ.) by area		area 2 -18% hosp. vs. area 1 in 1993/94 (H3N2)	only two seasons	[132]
Holland	1992/93-02/03	<50% pre-1996, 80% 2000-02	Rate ratios (RRs), 95% CIs: average AC mort. 1996-02 vs. 1992-95	Age, RSV, by high impact/H3N2 seasons	all RRs exclude null	confounding by temp.?	[64]
Canada	1997/98-03/04	excl. Ontario, 50% pre-2000, 70% 2000-03	RRs, 95% CIs: average AC, pneumonia and influenza hosp/A&E/GP rates 2000-03 vs. 1997-99	Age, sex, RSV, by H3N2 seasons, vaccine mismatch	all RRs exclude null	confounding by temp.?	[95]
France	1980/81-89/90	43% 1981, 72% 1989	plausible impact on seasonal mort. for range VE		7 - 697 deaths prevented per season	confounding by temp., RSV?	[79]

Simonsen *et al.* estimated the magnitude of the linear trend in $\log_{10}$ excess P&I and AC in the US from influenza A/H3N2-dominated seasons between 1980/81 and 2000/01, during which yearly vaccine coverage of the elderly increased from approximately 20% to 65%. [7] Excess mortality was calculated using a Serfling-type regression method described in section 2.4. After adjusting for ageing, the point estimate for the linear trend in excess P&I was -12% overall (-0.6% per year) with a lower bound of the 95% CI of -35%. The upper bound of the CI was not reported so it is not possible to say whether they had sufficient power to exclude the possibility of no change, or an increasing trend. A possible confounding effect of changes to ambient winter temperature between 1980 and 2001 was not assessed.

Rizzo *et al.* used the same method as Simonsen *et al.* to estimate the magnitude of the linear trend in $\log_{10}$ excess P&I and AC in H3N2-dominated seasons for two periods: the first when vaccine coverage was low (approx. 0 before 1980 to 8% in 1986) and the second when it was rising (10% in 1987 to 61% in 2001). [67] After adjusting for ageing, point estimates for linear trend in excess P&I were -146% overall (-9% per year) between 1970 and 1986 and +30% (+2% per year) between 1987 and 2001. No CIs were reported for these trends; the authors stated there was no evidence of a non-zero trend during the period of rising coverage which probably means the CI around the estimate of 30% increased excess P&I between 1987 and 2001 was wide and included the null. This would mean the analysis had little power to estimate the slope of the trend in this period.

Antunes *et al.* compared excess P&I in Sao Paulo, Brazil in two periods: before and after the introduction of free yearly influenza vaccination for the elderly. [131] In the period before free vaccination was introduced (in 1998), coverage was approximately 0; between 1998 and 2003 coverage ranged between 57% and 68%. Using Serfling-type least squares and ARIMA time series methods of calculating excess P&I, they estimated between 194 (47.4%) and 583 (88%) fewer excess P&I deaths in the period with vaccination compared to the period before vaccination. These estimates seem improbably high. Excess mortality for each season estimated using the two different methods (Serfling-type least squares and ARIMA time series modeling) produced very different estimates and different rankings of highest to lowest impact study seasons. This suggests there may have been problems in the analysis. It is possible that there is some con-

founding by antigenic drift in influenza A/H3N2 virus of the estimates of the decline of excess P&I with vaccination: there were 2 or 3 cluster transitions (depending on whether they were first isolated in the same influenza season as in the Northern hemisphere, or in the following influenza season) in the period of low vaccine coverage and only 1 during the period of high vaccine coverage. This might give the impression of an impact of vaccination on excess P&I which is in fact due to fewer antigenically drifted influenza A/H3N2 virus variants circulating in the vaccination period compared to the period before vaccination. The dominant circulating influenza type/subtype each influenza season was not stated. The report did not include CIs around estimated numbers of excess deaths prevented.

A study of two influenza seasons in Finland compared excess hospitalisation for pneumonia, influenza or circulatory disease in an area where people aged $\geq$ 65 yrs were only offered free vaccination if they were part of a risk group for complications of influenza with an area where all people aged $\geq$ 65 were offered free vaccination. [132] Taking excess pneumonia, influenza and circulatory hospitalisations together, the area that offered vaccination to all elderly, and that consequently achieved twice the vaccine coverage of the other area, had fewer excess hospitalisations in one of the two study seasons. The study season for which there was a clear difference in the excess hospitalisations between the two areas was dominated by influenza A/H3N2 virus; the season for which there was little difference in excess hospitalisations between the two areas was dominated by influenza B virus. For pneumonia or influenza hospitalisations alone, CIs for excess hospitalisations in the two study seasons and in the two localities overlapped considerably. The association between vaccination and excess hospitalisations was confined to circulatory hospitalisations in the higher impact influenza season. This study does not provide strong evidence for vaccine impact against circulatory hospitalisations because of its short duration (2 influenza seasons).

Jansen *et al.* calculated RRs comparing average excess AC mortality in the elderly in the Netherlands in the period of high vaccine use (coverage up to 80%, 1996/97 to 2002/03) to the period of low vaccine use (coverage <50%, 1992/93 to 1995/96). [64] Excess mortality was calculated using the rate difference method with either a 'peri-seasonal' or 'summer' baseline (described in section 2.4) and Serfling-type regression (similar to [7]). RRs for excess AC mortality attributable to RSV in high vs low coverage periods

were also calculated as a control. Influenza vaccine would not be expected to provide protection against RSV mortality. In a sensitivity analysis, the RRs were calculated based only on high impact influenza seasons (mostly H3N2-dominated seasons). This study provided robust evidence for vaccine impact against AC mortality: all RRs for excess mortality attributable to influenza excluded the null, associations were stronger when the analysis was restricted to high impact influenza seasons and, when the peri-seasonal baseline was used, vaccine impact against RSV-attributable excess mortality was not observed. The estimated impact of the increase in coverage of the elderly on mortality was 35-65%. There is less likely to be confounding by antigenic drift in this study than in the study by Antunes *et al.* since there were two cluster transitions in each of the low and high vaccine use periods. A possible confounding effect of different ambient winter temperatures in the two vaccine periods was not assessed. There could also be confounding by changing health status of the elderly over the study period, for example if levels of smoking had declined. This could have resulted in positive confounding of vaccine impact.

Using a similar design to Jansen *et al.*, Kwong *et al.* estimated RRs comparing excess AC and excess pneumonia or influenza hospitalisations, A&E visits and GP consultations for two periods: 1997-1999 (before the introduction of a universal influenza immunisation program (UIIP) in Ontario, Canada) and 2000-2003. UIIP offered free vaccination to all residents of Ontario regardless of age or risk group membership. Kwong *et al.* estimated influenza - attributable outcomes by regressing them on influenza indicators (see section 2.4.4). The purpose of the report was to demonstrate whether UIIP had led to greater impact in Ontario compared with impact of elderly/risk groups-only vaccination provided in the other Canadian provinces. The other provinces acted as a control for what the impact would have been in Ontario had UIIP not been in place. By looking only at the RRs for the other provinces (not Ontario), an estimate of the impact of the change in vaccine coverage of elderly $\geq$ 65 yrs of age from 1997-1999 (around 50%) to 2000-2003 (around 70%) can be gleaned. Controlled for age, sex, H3N2-dominance, RSV circulation and vaccine mismatch, all RRs exclude the null. The estimated impact of the increase in coverage of the elderly on mortality, hospitalisations and A&E visits is 50-70%, higher for GP consultations. These findings may be confounded by different ambient

winter temperatures or different health status of elderly in the two periods.

A different approach to estimating vaccine impact was that taken by Carrat and Valleron. [79] They calculated estimates of influenza - attributable mortality (from respiratory or cardiovascular causes, diabetes mellitus, chronic renal failure, cancer and other death rates) in the ≥ 75 age group prevented through vaccination of those ≥ 75. The number of deaths prevented through vaccination (d_a) was estimated for a range of plausible estimates of vaccine effectiveness (VE) from the literature, given the estimates of influenza - attributable mortality (d_o) and observed vaccine coverage in each influenza season (p). Influenza - attributable mortality was estimated by regressing the mortality rate registered to, for example, cardiovascular causes on the registered influenza mortality rate, with ARIMA errors (see section 2.4). Their formula for mortality prevented through vaccination was:

$$d_a = \frac{d_o VEp}{1 - VEp}$$

The formula was derived as follows (adapted from the appendix of [79]):

If

AR_u is the attack rate in the unvaccinated elderly
AR_v is the attack rate in the vaccinated elderly
n_u is the number of unvaccinated elderly
n_v is the number of vaccinated elderly

given

$$VE = 1 - \frac{AR_v}{AR_u}$$

then

$$d_o = \frac{(AR_u n_u + AR_v n_v)}{(n_u + n_v)} = \frac{AR_u(n_u + (1 - VE)n_v)}{(n_u + n_v)}$$

Replacing $n_v/n_u = p/1 - p$ and simplifying gives

$$d_o = AR_u(1 - VEp)$$

Assuming that AR_u does not depend on vaccine coverage, AR_u is the estimate of the death rate in the absence of vaccination. (Note that AR_u is unobserved.)
Vaccine impact (d_a) is then

$$d_a = AR_u - d_o = \frac{d_o}{(1 - VEp)} - d_o = \frac{d_o VEp}{(1 - VEp)}$$

The difference between the (unobserved) AR_u and the observed influenza-attributable mortality d_o is attributed to vaccination. Assuming that the attack rate in the unvaccinated does not depend on vaccine coverage means that this formula is estimating vaccine impact due to direct effects of vaccination only. The authors did not report mortality prevented in each season. They estimated that between approximately 7 and 700 influenza - attributable deaths may have been prevented in the influenza seasons studied, depending on the true VE and coverage achieved in that season. This method could be refined in future by using influenza season-specific estimates of VE against mortality, if available. This would produce estimates of mortality prevented in each of a number of influenza seasons controlled for variability in VE due to vaccine mismatch or lower impact influenza seasons. Carrat and Valleron's application of the method did not allow for confounding by other factors, such as ambient temperature or RSV epidemics, which, if they overlapped with the period of influenza circulation, might result in higher estimates of influenza - attributable mortality and consequently estimates of impact that are lower than in reality.

Estimated vaccine impact against similar outcomes may vary between settings because of setting-specific factors affecting estimates of the relative impact of influenza seasons (e.g. diagnostic practices, access to care, ambient winter temperatures). On balance, the studies reviewed in this section suggest that there may be an impact of increasing vaccination of the elderly on morbidity and mortality in the elderly. The reports by Simonsen *et al.*

and Rizzo *et al.* suggest vaccine impact may be difficult to detect using linear regression because of low power. [7, 67] Results presented by Jansen *et al.* and Kwong *et al.* would suggest that for an increase in coverage of the elderly of approximately 20-30%, a decline in influenza - attributable mortality of 35-70% could be expected. [64, 95] Despite careful control for confounding, it is unlikely that these estimated declines in mortality are entirely attributable to vaccination. Viboud and Miller showed similar declines in mortality in the Northern US despite stagnant vaccine coverage in the US over this period. [133] Carrat and Valleron's model could provide a method of estimating vaccine impact each season based on season-specific estimates of vaccine effectiveness, excess mortality (or morbidity) and vaccine distribution, assuming only direct effects of vaccinating the elderly. [79]

2.10. Rationale for overall aims and objectives

In this final section, the rationale for the work described in the thesis is given. As outlined in the preceding literature review, the average effect of CT seasons on influenza - attributable morbidity and mortality in England & Wales (or, indeed, globally) has not been quantified. The impact of the increase in yearly vaccine coverage of the $\geq$ 65 age group in England & Wales from 24% in 1989/90 to 71% in 2004/05 on morbidity or mortality in the elderly is also unknown.

The effect of CT seasons and the impact of vaccination can both be estimated by expressing influenza - attributable mortality as dependent on CT seasons or vaccine coverage in a regression model. A limitation of the regression models most frequently used to model influenza - attributable morbidity and mortality, and determinants of these, is the need to externally designate expected morbidity or mortality if influenza virus had not been circulating (where morbidity or mortality in excess of this expected morbidity or mortality is attributable to influenza). The various ways that this is routinely done leave the possibility of counting unrelated deaths or consultations as influenza - attributable (lowering the specificity of influenza - attributable mortality for influenza) or of designating truly influenza - attributable deaths or consultations as baseline and thus unrelated to influenza (potentially biasing downwards estimates of the effect of determinants of influenza - attributable morbidity or mortality by obscuring the full extent of

the variability in influenza - attributable morbidity and mortality between influenza seasons). Apportioning deaths or consultations to influenza and to other aetiologies using multiple linear regression relies heavily on laboratory data that are of limited value for informing the relative impact of influenza seasons over long time periods because of increased amounts of testing over time. In the work described in the thesis, latent variable time series models were fitted in which the designation of baseline from influenza - attributable morbidity and mortality was made by the model. Laboratory data for influenza A were used to inform the timing of influenza seasons by fitting latent variable time series models jointly to morbidity/mortality and laboratory data (morbidity/mortality and laboratory data were simultaneously modeled as outcome variables and were regressed on indicators for seasonality, trend and data artifacts - see chapter 3). Laboratory data did not inform the relative impact of influenza seasons, only their timing.

The latent variable models described in the thesis defined influenza - attributable morbidity or mortality using a simple random effect mean shift (the difference between baseline and influenza - attributable morbidity or mortality in models with an identity link or the rate ratio comparing baseline to influenza - attributable morbidity or mortality rates in models with a log-link (see section 3.5.2)). Attempts were then made to estimate the average effect of CT seasons on influenza - attributable morbidity and mortality by expressing the influenza season-specific random effect mean shift as dependent on a binary variable designating seasons as being CT or intracluster (non-CT) seasons. The variability in the size of cluster transitions was allowed for in a supplementary analysis where the random effect mean shift was expressed as dependent on a quantitative variable for the size of cluster transitions (0 for intracluster seasons). In separate analyses, attempts were made to estimate the impact of each unit increase in the vaccine coverage of the ≥ 65 age group by expressing the random effect mean shift as dependent on a quantitative variable for vaccine coverage each season.

3. Methods

3.1. Aim of this chapter

In this chapter, general methods used in the thesis are introduced. Detail of specific models is provided in results chapters.

3.1.1. Objectives of this chapter

1. To describe data sets analyzed, data management undertaken and relevant epidemiology of influenza
2. To describe frequentist generalised linear models used to estimate the shape of long-term trend in P&I and ILI
3. To describe Bayesian Markov chain Monte Carlo methods to fit two-state hidden Markov models to P&I, ILI and laboratory reports for influenza A

3.2. Data sets

Influenza incidence must be estimated indirectly because laboratory confirmation is seldom done. In the next five subsections the data analyzed in the thesis are discussed. These data are Royal College of General Practitioner sentinel general practitioner consultations for influenza-like-illness (section 3.2.1), all registered deaths coded to underlying pneumonia or influenza from the Office for National Statistics (section 3.2.2), laboratory reports for influenza A and B from the Health Protection Agency Centre for Infections (section 3.2.3), vaccine coverage of the 65+ age group in each influenza season from 1989/90 to 2004/05 from the literature (3.2.4) and cluster transitions in the evolution of influenza A/H3N2 virus from the literature (3.2.5). Finally, lags between timing of 'aberrant' periods in P&I, ILI and laboratory data for a given age group are discussed in section 3.2.6.

3.2.1. Royal College of General Practitioners influenza-like-illness consultations

The RCGP has collected the weekly number of GP consultations for new episodes of ILI by age group (0-4, 5-14, 15-44, 45-64, 65+) and by sex through their WRS since 1967. The WRS represented morbidity statistics for a registered population of 200,000 people in England & Wales in 1967 and over 600,000 people in 2005. [122, 134] Reports of ILI are not based on a case definition but rather on a GP diagnosis made using suggested diagnostic guidelines provided by the RCGP. New episodes are defined as those occurring 21 days after a previous episode, but not all GPs adhere to this rule. [135] Reporting physicians are encouraged to use diagnostic Read codes (as opposed to symptom codes which are what patients present with or what they say they feel/have); this is thought to encourage the reporting GP to make a diagnosis. The RCGP provided weekly counts of ILI consultations by age group from 1967-2005, and the weekly registered population, for use in the work described in the thesis.

Validation

ILI data have been validated for influenzal illness incidence in several ways. [6, 136,137] Comparing ILI data with virological data shows that increased clinical incidence does not occur in the absence of increased laboratory reports for influenza. [6] Patterns in ILI incidence have been shown to agree with patterns in respiratory hospitalisation and all-cause and respiratory mortality. [136, 137]

Representativeness

The population registered with sentinel physicians is representative of the population of England & Wales in terms of age, sex and social deprivation. [138–141] RCGP state the practices were representative of the National population by age and sex from 1966 to 1983. [138] The report *Morbidity Statistics from General Practice: fourth national study 1991-92* (MSGP4), where most participating practices were also contributing to the WRS, suggested MSGP4 participating practices, and hence most WRS practices, were representative of population of England & Wales in terms of age, sex, marital status, housing tenure and the proportion who smoked. [142, 143] In the

MSGP4 report, the General Household Survey was used to indicate the demographic profile for England & Wales. Particular ethnic groups, the lowest socioeconomic status (SES) band, the South of England and metropolitan areas were under represented in MSGP4 participating practices. Recent expansion of the WRS improved in particular representation of the lower SES band and the South of England in the WRS. [141]

Limitations

There is no case definition for ILI in the WRS and diagnostic guidelines have only been given to the GPs since 1991. [144] Specificity of the diagnosis of ILI varies during the year because laboratory surveillance suggesting influenza virus is circulating in the community is publicised to participating GPs. This leads to a particular clinical presentation being more likely to be diagnosed as ILI when influenza virus is known to be circulating in the community than during other times of the year. Recall that laboratory surveillance underestimates summer circulation of influenza virus. There are no firm rules on the length of an episode. Reporting delays are common around holiday periods. [53]

ILI consultation data have low sensitivity for influenza. A full picture of the impact of influenza on primary care includes consultations for acute bronchitis, otitis media and asthma. [3] ILI have been modeled for the thesis because of their relative specificity for influenza.

Data management

Between 1994 and 1998, the RCGP received both electronic and paper-based weekly returns from participating practices. From 1999, all weekly returns were electronic (see section 3.2.1). Discussions with RCGP staff confirmed that all data (reported electronically or on paper) had been provided for this analysis.

Artifacts

Several changes in the way the WRS operated over the study period were noted and controlled for in the analysis using dummy variables (table 3.1). In the period 1988-90, several practices changed from paper to computer-based recording and there was subsequent evidence of consistently higher

level of ILI reported from computerised practices. [145] From the 1993-94 influenza season, 'epidemic influenza' (a more severe illness) and 'influenza-like illness' (a less severe illness) were amalgamated for presentation of rates of ILI. [146] Though the data analyzed here are 'epidemic influenza' plus 'influenza-like-illness' combined for the whole time series, a dummy variable for this change was included in case the change affected the data. In the 1999-2000 influenza season, reporting by practices to the RCGP became fully-automated with a consequent decline in the number of reporting practices that has since rebounded. [144]

Table 3.1.: **Known artifacts in P&I and ILI.**

Dataset	Year	Change
ILI	1988	transition from paper-based to electronic reporting began
	1993	'epidemic influenza' and 'influenza-like-illness' now combined
	1999	full automation of reporting from practices to RCGP
P&I	1979	change of ICD version (8 to 9)
	1984	change of how rule 3 for underlying cause of death applied
	1993	change of software for coding underlying cause of death
	2001	change of ICD (9 to 10) and new interpretation of rule 3

3.2.2. Office for National Statistics mortality data

The Office for National Statistics (ONS) has electronically archived underlying and contributing causes of death abstracted from death certificates since 1970. [147] ONS provided daily numbers of deaths registered to P&I from 1970 to 2004 by age group (<1, 1-4, 5-84 in 5-year age groups, 85+). P&I was defined using the following International Classification of Disease (ICD) codes: ICD8: 470-474, 480-483, 485-486, ICD9: 480-487 and ICD10: J10-J18. Influenza is rarely mentioned on death certificates as an underlying or contributing cause of death. [47] An average of 0.3% of deaths were registered to underlying influenza between 1993-2003 (0.6% in 1999/00 influenza

season). [148] For 80% of death certificates where influenza is mentioned, it is chosen as underlying cause but it is very rarely diagnosed. It is likely that many of those who die as a result of influenza infection die from pneumonia. [45] P&I was analyzed as a proxy for deaths from influenza because P&I has greater specificity than all-cause or respiratory deaths.

Validation

P&I is a valid proxy for deaths from influenza. P&I is positively associated with circulation of influenza viruses. [41,149] P&I has been shown to reliably and specifically indicate timing and relative size of influenza seasons in the US. [149–151]

Representativeness

The data provided by ONS are all deaths registered to underlying P&I in England & Wales.

Limitations

Death registrations, not occurrences, have been analyzed; there are rarely more than 2-3 days between death and registration except around public holidays. [47] In the thesis, splines used to model long-term trend smooth out blips from public holidays. As for ILI, P&I has variable specificity for influenza since physicians are more likely to diagnose influenza if laboratory reports for influenza suggest that influenza virus is circulating in the community at the time of the death. P&I has low sensitivity for influenza and the full picture of deaths from influenza includes deaths from other respiratory causes (e.g. bronchitis) as well as cardiovascular diseases. [69,152] P&I was modeled because it has higher specificity for influenza than all-cause, respiratory or respiratory plus circulatory deaths and therefore is best suited to estimating the relative impact of influenza seasons.

Data management

Daily counts of P&I were collapsed into age groups to match RCGP data (0-4, 5-14, 15-44, 45-64, $\geq$ 65). Daily age group-specific counts of P&I were collapsed into weekly age group-specific counts in Stata version 9. [153]

Dates contributing to each week were reconciled with the ILI time series: weeks ran Thursday to Wednesday from January 1967 to August 1969, Wednesday to Tuesday from August 1969 to August 1991 and Monday to Sunday from then on. August 13th 1969 was counted twice and there was one 5-day week (August 28 - September 1 1991).

Artifacts

Several changes to the coding of underlying cause of death were adjusted for using dummy variables (table 3.1). In 1984, ONS introduced broader interpretation of rule 3 for coding underlying cause of death (where a condition in part I or II of the death certificate could take precedence over the cause of death selected using other rules if the cause was a direct consequence of the condition in part I or II). [147] This led to an abrupt fall in deaths registered to underlying pneumonia. In 1993, ONS adopted an automated system for coding cause of death which narrowed the interpretation of rule 3 and approximately reversed the change adopted in 1984. [154] With the change from ICD 9 to 10 in 2000, respiratory deaths fell by approximately 22% while deaths coded to pneumonia fell by 38%. [148]

3.2.3. Laboratory surveillance

The HPA CfI provided individual records from LabBase2 of all laboratory-confirmed influenza A infections, based on virus isolation and PCR, reported voluntarily by National Health Service (NHS) and HPA laboratories in England & Wales between 1975 to 2005. [10] Records included individuals' age, sex, the site of influenza virus isolation and the earliest specimen date.

Representativeness

Laboratory reporting is voluntary so it is difficult to assess the consistency of geographic representativeness of laboratory reports. Until 1993 the majority of influenza laboratory tests were for hospitalised patients. In 1993 two sentinel swabbing studies were introduced in general practice, [33, 34] one of which deposits positive results in LabBase2. [33]

Limitations

Diagnostic techniques are not standardised so sensitivity and specificity may vary between laboratories reporting positive influenza specimens to HPA CfI. [155] The earliest specimen date is the date of report to the HPA CfI from 1975-1988, when testing was done in batches approximately weekly, and the date of specimen collection after 1988. The date of infection is rarely known. [156] The date of specimen collection was missing in approximately 3% of reports in 1992; [10] for these reports the earliest specimen date is the date of report. Approximately 10% of laboratory reports have missing age or sex. Before the introduction of the LabBase2 electronic database (in 1989) approximately 20% of reports had missing specimen type.

Data management

From individual laboratory reports, age group-specific (0-4, 5-14, 15-44, 45-64, $\geq$ 65 years) weekly counts were created. Weeks were reconciled to match dates in the ILI time series. Approximately 10% of laboratory reports were excluded from the analysis because of missing age or sex or because site of isolation was gastrointestinal, cerebrospinal or genitourinary. Weekly time series were plotted overall and by age group to check for unknown artifacts.

Artifacts

HPA CfI adopted the electronic database LabBase2 in 1989 for storage of laboratory reports of influenza. [10] In the 1993/94 season two sentinel GP swabbing studies began where GPs were asked to swab a number of ILI patients each week during winter. [33, 34] The Virus Reference Unit(VRU)/RCGP collaborative study data are not included in LabBase2 [34] but HPA sentinel surveillance data typically are included. [33] These artifacts were controlled for in all analyses by using dummy variables.

3.2.4. Vaccine coverage in the over 65s

Vaccine coverage of those $\geq$ 65 years of age in England & Wales during the influenza seasons 1989/90 to 2004/05 was derived from published sources. [119, 120] Coverage between 1989/90 and 2003/04 was estimated using the General Practice Research Database (GPRD). [120] The GPRD

holds patient electronic records for 350 practices in the UK and links prescribing information to outcome and exposure events (e.g. vaccination). Joseph *et al.* estimated vaccine coverage, by age group, sex and risk status categories, for each influenza year (1 July to 30 June). [118, 120] Risk status was defined using clinical codes - Oxford Medical Information Systems (OXMIS) and Read codes - for Department of Health defined risk groups for influenza complications. Patients with no recorded risk were defined as low risk. All people $\geq$ 75 were defined as high risk from 1998/99 onwards and all people $\geq$ 65 were defined as high risk from 2000/01 onwards to reflect government vaccination policy. Coverage from 2004/05 was ascertained from Department of Health surveillance of vaccine provision in general practice. [119]

Validation

GPRD has good validity for prescriptions which approximates validity for exposure events like vaccination. [157] The estimate of coverage for 2004/05 from Department of Health surveillance of vaccine uptake in general practices has not been validated.

Representativeness

GPRD is representative of the UK population in terms of age and sex. [113] Representativeness of Department of Health surveillance of vaccine uptake in general practices is unknown; the aim of the scheme is to ascertain all influenza vaccinations administered in general practice.

Limitations

Both sources of vaccine coverage information are likely to underestimate coverage because of vaccines administered outside of general practice.

Data management

Estimates of coverage of those $\geq$ 65 pre-2000/01 (when policy changed to include all people aged 65-69 years in the risk group recommended vaccination) were calculated by summing numbers vaccinated in 65-74 and $\geq$ 75 age groups and dividing by the number of people aged 65-74 and $\geq$ 75. [120]

3.2.5. Influenza A/H3N2 virus cluster transitions

Seasons with CTs, and the antigenic distance (AgD) between clusters, were taken from Smith *et al.* and are shown in table 3.2. [13]

Table 3.2.: **Influenza seasons in England & Wales, 1968/69-2004/05: dominant variant, H3N2 CTs, size of CTs in terms of antigenic units, and vaccine mismatch.** Bolding means H3N2 dominant and vaccine variants are from different clusters.

Season	Dominant variant[ref]	CT [13]	CT size (Ag units) [13]	H3N2 vaccine cluster[ref]
1968/69	H3N2 [158]			
1969/70	H3N2 [158]			
1970/71	B [158]			
1971/72	H3N2 [158]			
1972/73	H3N2 [158]	HK68-EN72	3.4	
1973/74	B/H3N2 [158]			
1974/75	H3N2 [158]			
1975/76	H3N2/B [158]	EN72-VI75	4.4	
1976/77	H3N2 [159]			
1977/78	H3N2/H1N1 [158]	VI75-TX77	3.4	
1978/79	B [158]			
1979/80	H3N2 [158]	TX77-BA79	3.3	
1980/81	H1N1/H3N2 [158]			
1981/82	B/H3N2 [160]			
1982/83	H3N2 [160]			
1983/84	H1N1/B [160]			
1984/85	H3N2/B [160]			
1985/86	B [161][1]			
1986/87	H1N1 [161][1]			
1987/88	H3N2/H1N1 [6]	BA79-SI87	4.9	
1988/89	H1N1/H3N2 [6]			
1989/90	H3N2 [6]	SI87-BE89	4.6	**SI87** [162]
1990/91	B [6]			BE89 [163]
1991/92	H3N2 [6]			BE89 [164]
1992/93	B/H1N1 [6]	BE89-BE92	7.8	**BE89** [165]
1993/94	H3N2 [6]			BE92 [166]
1994/95	B [6]			BE92 [167]
1995/96	H3N2 [6]	BE92-WU95	4.6	**BE92** [168]
1996/97	H3N2 [6]			WU95 [169]
1997/98	H3N2/H1N1 [170]	WU95-SY97	4.7	**WU95** [170]
1998/99	H3N2/B [171]			SY97 [171]
1999/00	H3N2 [172]			SY97 [172]
2000/01	B/H1N1 [173]			SY97 [173]
2001/02	H3N2/H1N2 [174]			SY97 [174]
2002/03	B/H3N2 [175]	SY97-FU02	3.5	**SY97** [175]
2003/04	H3N2 [176]			**SY97** [176]
2004/05	H3N2 [177]			FU02 [177]

[1] based on Netherlands

Validation

The accuracy of the antigenic map was determined by selecting antigen-antiserum pairs whose HI values were not included in the creation of the map, predicting the HI value of those strains from distances in the map, and then testing the HI value for this pair in the laboratory. The correlation between HI distance inferred from the map and laboratory-tested distance was 0.81. [13]

Representativeness

Ninety-four of 273 isolates used to create the antigenic map were from the Netherlands; the remainder from elsewhere. The same type/subtype was dominant in both the Netherlands and in England & Wales in each season between 1987/88 and 1996/97. [6] Also, molecular epidemiology studies of the global spread of influenza A/H3N2 virus suggest that global antigenic drift patterns may dwarf local patterns, providing some reassurance that the seasons in which CTs were first isolated in the data with a Dutch bias reflect the seasons the CTs were first isolated in England & Wales. [101,178]

Limitations

The lack of global representativeness of the HI data.

Data management

CTs were coded as being dominant from the season of first isolation, or from the first influenza A/H3N2 virus - dominated season after first isolation if the CT was first isolated in a season dominated by influenza A/H1N1 or B viruses. This was done to allow that a CT first isolated during a season not dominated by influenza A/H3N2 would be unlikely to affect incidence until it was the dominant variant in circulation. The dominant variant from each season was taken from the literature (table 3.2). Influenza A/H3N2 virus was considered dominant if it alone dominated, or if it was co-dominant alongside influenza A/H1N1 or B virus.

3.2.6. Lags between timing of 'aberrant' periods in P&I, ILI and laboratory reports

If an individual were captured by all three outcome data sets (P&I, ILI and laboratory data), they would be expected to be ascertained first in ILI, concurrently or secondly by laboratory reports (historically most laboratory testing for influenza occurred in hospitals while after 1993 the laboratory data are a mixture of hospital and GP results) and thirdly upon death registration which occurs typically within 2-3 days of death. [47] There is, however, incomplete overlap between people in the 3 data sets analyzed because of different severities of illness (a mild illness that necessitated seeking care might only be ascertained by ILI) and different underlying conditions (an asthmatic might bypass the GP surgery and be ascertained first when admitted to hospital with exacerbation of their asthma and subsequently tested for influenza while in hospital). Typically, those who die from influenza are elderly. There are also lags inherent in the data themselves: the lag between death and death registration and between date of infection, date of specimen collection and date of report in laboratory data are examples. Changes to policy (e.g. the *Path of Least Resistance* report in 1998 released to help curb antibiotic prescribing for respiratory illness in GP surgeries [179]) and to surveillance systems (e.g. introduction in 1993 of two sentinel swabbing schemes in GP practices to test a proportion of people with ILI for influenza [33,34]) means that the mixture of people, in terms of severity of disease and underlying illness, captured by the three data sets has changed over time. This issue is explored in more detail in chapter 6.

3.3. Epidemiology of influenza

The epidemiology of influenza dictates methods of analyzing influenza incidence data. First, influenza virus is infectious and cases infect others in a chain of transmission. This leads to non-independence of counts from one week to the next (called autocorrelation); space-time clustering of cases leads to overdispersion of incidence data relative to the Poisson distribution. Overdispersion means that the variance of the data is greater than their mean. The Poisson distribution is defined by a single parameter which

is both mean and variance. Second, recall from section 2.2 that influenza incidence in temperate climates is strongly seasonal. The consequent correlation between observations approximately one year apart must be taken into account in an analysis. Finally, recall also from section 2.2 that 'normal' seasonal influenza morbidity and mortality is punctuated with 'aberrant' morbidity and mortality most years, and infrequently by epidemics. This pattern of incidence contributes to influenza morbidity and mortality data being overdispersed relative to the Poisson distribution. Influenza can be modeled using transmission dynamic or statistical models. There is a long history of statistical modeling of influenza (summarised in section 2.4). Definition of morbidity and mortality that would have occurred had influenza virus not been circulating is a key feature of the most commonly used models. [60] The latent variable time series models accounting for seasonality that have been fitted for the thesis address each of these aspects of the epidemiology of influenza.

3.4. Generalised linear modeling of baseline trend in P&I and ILI

In order to calibrate fitting of long-term trend in the latent variable time series models (section 3.5.2), generalised linear models (GLM) with a log-link were fitted to P&I and ILI by age group, accounting for seasonality, long-term trend and artifacts. The log-link allowed inclusion of a population offset in the linear predictor to account for the changing size of the population at risk over time. [180] Different long term trends by age group were tested for. Various methods of fitting long-term trend were evaluated.

3.4.1. Generalised linear models

Counts of disease are often modeled using the Poisson distribution. [180] The Poisson mean equals its variance. Because influenza incidence data are overdispersed relative to the Poisson distribution (section 3.3), negative binomial instead of Poisson GLMs were fitted. Negative binomial GLMs allow that the variance of a data set exceeds its mean. [181] The negative binomial model with a log-link has the following form

$$
\begin{aligned}
Y_t &\sim \text{Poisson}(\mu_t\theta) \\
\log(\mu_t) &= \alpha + \beta t \\
\theta &\sim \text{Gamma(a, b)}
\end{aligned}
$$

Y_t is the observed count in week t, which is a Poisson random variable with mean $\mu_t\theta$; θ is a Gamma-distributed dispersion parameter. The parameters of the Gamma distribution that θ follows (a and b) control overdispersion relative to the Poisson distribution. Values of θ greater than 1 indicate the data are overdispersed relative to the Poisson model. The $\log(\mu_t)$ is modeled as dependent on α, the intercept, and β the linear trend.

3.4.2. Excluding epidemic weeks

Between 1 and 25% of the highest counts, in each age group of each data set, were excluded from the fitting of the negative binomial models in an attempt to isolate 'normal', or baseline, from 'aberrant', or influenza-attributable, incidence. [61]

3.4.3. Changing population size

To be able to assess trends in morbidity or mortality over time, the changing size of the population at risk should be adjusted for. Crude trends in age-specific rates are meaningful for assessing age-specific trends. Rates over time for all ages combined should be adjusted for the changing age-distribution of the population over time before presence of a trend is tested. Direct age-standardisation assumes that trends are similar in each age group, an assumption that should be checked (chapter 4). [182] If the rate of antigenic drift in influenza A/H3N2 virus is slowing, for example, incidence of infection over time in the younger ages would remain approximately constant (since people are born susceptible, ignoring maternal antibodies), while rates of infection in middle aged people would decline (because of cross immunity and declining mutability).

The changing size of the population at risk was modeled as a population offset, $\log(N_t)$, as below:

$$\log(\mu_t) = \log(N_t) + \alpha + \beta t$$

For ILI data, the population offset was the number of registered patients

on the middle day of each week (i.e. Thursday for a Monday-Sunday week) and for P&I data it was the relevant census population of England & Wales or an inter census estimate. [183] Rearranging the formula gives

$$\log(\mu_t/N_t) = \alpha + \beta t$$

showing the log-link model including a population offset is equivalent to modeling the log rate.

3.4.4. Seasonality

Seasonality in the influenza incidence data was captured with one Fourier term (i.e. one sine and one cosine term) as below

$$\log(\mu_t) = \log(N_t) + \alpha + \beta t + \beta_1 \sin\frac{2\pi t}{52.2} + \beta_2 \cos\frac{2\pi t}{52.2}$$

The period is 52.2 weeks, not 52 weeks, to account for not every year having exactly 52 weeks. This method of capturing yearly seasonality in influenza morbidity and mortality data was introduced by Serfling [70] and has been used by many authors (e.g. [60]). This method of modeling yearly seasonality in data is more parsimonious than, for example, including a term for month (2 parameters vs 12 parameters) [52,184] or modeling trend and seasonality simultaneously with a spline with several degrees of freedom (df) for each year. [49,185]

3.4.5. Long-term trend

Long-term trend was modeled using linear (βt) or quadratic (βt^2) terms or cubic splines with up to 20 df. The model capturing long-term trend with cubic splines is shown below:

$$\log(\mu_t) = \log(N_t) + \alpha + C(t, \varphi) + \beta_1 \sin\frac{2\pi t}{52.2} + \beta_2 \cos\frac{2\pi t}{52.2}$$

The number of df is φ. Cubic splines flexibly model trend with a number of cubic curves smoothly connected at knot points. [186] Splines captured variability in the baseline in addition to that explained by seasonality and by artifacts (see next section). Best-fitting models were chosen by comparing Akaike's Information Criteria (AIC) to balance goodness of fit and parsimony (section 3.4.8).

3.4.6. Artifacts

Step changes in the long-term trend due to known artifacts (table 3.1) were fitted via categorical dummy variables ('artifacts') as below:

$$\log(\mu_t) = \log(N_t) + \alpha + C(t, \varphi) + \beta_1 \sin\frac{2\pi t}{52.2} + \beta_2 \cos\frac{2\pi t}{52.2} + \beta_3 \text{artifacts}$$

3.4.7. Inference

The likelihood function is the probability of the data conditional on the parameters (which are fixed). [180] For the negative binomial model, the log likelihood (LL) is

$$LL = \sum_{t=1}^{n} \left(Y_t \log(\mu_t) - \log(Y!) - \frac{1}{\theta}\log(1 + \mu_t\theta) + \log\left(\frac{\Gamma(Y_t + 1/\theta)}{\Gamma(1/\theta)(1/\theta + \mu)^{Y_t}}\right)\right)$$

where Y_t is the observed count, μ_t is the predicted count and θ is the dispersion parameter.

GLMs were fitted by maximum likelihood, which means that the value of each parameter was estimated such that the value of the likelihood function was maximised. The data are analyzed in isolation of prior knowledge (from the literature, for example) as to the value of model parameters. Estimates of the uncertainty about values of parameters of interest - confidence intervals (CI) - are expressed as the interval that includes the true value in e.g. 95 or 99% of (hypothetical) repeated samples, if the experiment were repeated many times with all parameters constant.

3.4.8. Model selection

Candidate models with linear or quadratic terms to model trend or with cubic splines with different df to model trend, with and without age group - trend interaction terms, were compared using AIC. [187] AIC are goodness of fit statistics calculated as

$$AIC = -2LL + 2npar$$

where npar is the number of parameters in the model. Lower AIC indicate

better model fit. AIC balance precision, i.e. the highest log likelihood value, with complexity of the model in terms of the number of parameters. AIC allow comparison of non-nested models as long as the models are fitted to the same data.

3.4.9. Rationale for hidden Markov models

GLM modeling of influenza-related morbidity and mortality is hindered by problems of autocorrelated residuals and overdispersion not adequately addressed with the negative binomial model (see section 4.6). These issues indicate that a type of latent variable model called a hidden Markov model (HMM) may be useful. It can be shown that HMMs are autocorrelated and overdispersed. [188] Serial correlation arises from the underlying Markov chain and overdispersion arises because observations are modeled as arising from one of several marginal distributions, each associated with a different hidden state. [188, 189] In the context of influenza, it is natural to think of non-laboratory confirmed time series related to influenza as having arisen from two marginal distributions: one 'normal' and one 'aberrant'. The remainder of this chapter describes methods for fitting two-state HMMs using Markov chain Monte Carlo (MCMC) in OpenBUGS.

3.5. Bayesian Markov chain Monte Carlo fitting of hidden Markov models

HMMs were fitted in order to allow modeling of 'normal' and 'aberrant' incidence simultaneously where the differentiation between 'normal' and 'aberrant' incidence is automated. Univariate two-state HMMs were fitted to weekly P&I counts by age group and weekly ILI counts by age group (chapter 5). Bivariate (P&I + laboratory reports for influenza A in one set and ILI + laboratory reports for influenza A in another) and trivariate (P&I, ILI and laboratory reports in the same model) models were fitted in chapter 6.

Figure 3.1 shows a schematic of a HMM. t denotes a time interval (in this case one week) and arrows denote conditional dependencies. The observed counts are independent conditional on the unobserved states (details in section 3.5.2).

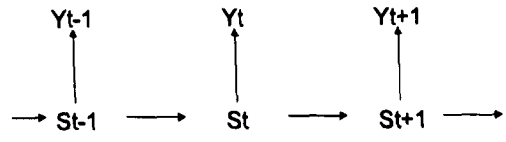


Figure 3.1.: **Schematic of a HMM.**

3.5.1. Bayesian Inference

In chapters 5 to 7, Bayesian inference is used in order to incorporate uncertainty in the differentiation of 'aberrant' from 'normal' incidence in HMM model fits and to allow prior information as to the effect of cluster transitions and the impact of vaccination on mean shifts to be incorporated (chapter 7).

Bayesian inference is based on the posterior distribution. Let Y denote the observed data and β denote model parameters. Bayes theorem states that

$$P(\beta|Y) = \frac{P(\beta)P(Y|\beta)}{\int P(\beta)P(Y|\beta)d\beta}$$

i.e. that the probability of the parameters, given the data ($P(\beta|Y)$) is equal to the product of the likelihood $P(Y|\beta)$ and the probability of the parameters $P(\beta)$, divided by a normalisation constant ($\int P(\beta)P(Y|\beta)d\beta$). The probability of a parameter is also called the prior for that parameter. Priors are the value of key parameters, with uncertainty, before looking at the observed data that are being analyzed. Full Bayesian analyses thus put the data being analyzed in the context of what is already known. In making inference using the posterior distribution, estimates of uncertainty around parameters of interest - credible intervals (CrI) - are 95 or 99% cut-points of the posterior distribution that contain 95 or 99% of the sampled values for that parameter.

3.5.2. Hidden Markov models

In HMMs, observations are independent conditional on an unobserved state variable. HMMs have been applied to modeling ILI data from France, [81] P&I data from the US, [82] hospital infection data in England & Wales, [89]

and have been used extensively in other disciplines (e.g. [189]). Below is a univariate two-state Poisson HMM with a log-link. The difference between the 'normal' and 'aberrant' incidence is modeled as random effect mean shift for each influenza season (α_1[flu season]).

$$
\begin{aligned}
Y_t &\sim \text{Poisson}(\mu_t) \\
\log(\mu_t)|\ S_t = 1 &= \log(N_t) + \alpha_0 + \beta t \\
\log(\mu_t)|\ S_t = 2 &= \log(N_t) + \alpha_0 + \alpha_1[\text{flu season}] + \beta t \\
S_t|\ S_{t-1} &\sim \text{Bernoulli}(\delta)
\end{aligned}
$$

Y_t is the observed number of deaths in week t,
μ_t is the mean of the Poisson distribution from which Y_t is drawn,
N_t is the population offset,
α_0 is the intercept,
α_1[flu season] is the mean shift estimated when the model is in state 2,
βt is the linear trend,
and S_t is the state variable sampled from a Bernoulli distribution with probability δ, a two-by-two matrix of probabilities of moving between states at time t given the state of the model at time t-1.

Each week, Y ILI or P&I are observed. Y is assumed to be a realisation of a random process, $Y = (Y_t; t = 1, n)$, where each Y_t is associated with an unobserved random variable S_t. S_t determines the conditional distribution of Y_t. If $S_t = j$, then the conditional distribution of Y_t has density $f(j) = (Y_t; m(j))$ where $f(j)$ belongs to a given parameterised family and $m(j)$ are parameters to be estimated. In this case S_t follows a two-state homogeneous Markov chain of order 1, meaning the probability of moving between states does not change with time (is homogeneous) and the process recalls only the state at time t-1 (is a Markov chain of order 1, also called a random walk). The probability of moving from state i to state j at time

t is expressed as $\delta(ij) = P(S_t = j|S_{t-1} = i), i, j = 1, 2$. Each state has a Poisson observation model associated with it. Covariates are incorporated by expressing parameters (e.g. β) as dependent on the covariates.

As introduced in section 3.4.9, serial correlation is generated in the HMM from the underlying Markov chain. [188] Overdispersion of the Poisson HMM relative to the Poisson GLM comes about because observations are modeled as arising from one of several marginal distributions, each associated with a different hidden state. [188, 189] Counts predicted if the 'aberrant' state is drawn are higher than if the 'normal' state is drawn by a random effect mean shift estimated for each influenza season (α_1[flu season]): the mean shift is multiplicative, on the original scale, in the Poisson HMM with a log-link and additive in the Poisson HMM with an identity-link.

3.5.3. Markov chain Monte Carlo

In a HMM, inference about values of parameters of interest cannot readily be made by directly maximising the likelihood of the data and the parameters for every possible sequence of states because of computational intractability. [81,89] Algorithms like the Expectation - Maximisation (EM) algorithm, or similar recursive methods, can be used. These algorithms work by first estimating the conditional expectation of indicator variables for the two states (the expectation step) and applying the values of the 2 indicator variables to the complete data LL to estimate the vector of transition probabilities and all parameter values (the maximisation step). [81, 89] In order to additionally incorporate prior information as to the value of different parameters of interest, HMMs were instead fitted in a Bayesian framework where sampling from the joint posterior distribution of the likelihood and priors was done using MCMC.

MCMC is a way of fitting models via stochastic simulation, usually in the Bayesian framework. It is an alternative to the EM algorithm for repeated sampling from a complex surface like a HMM. A simple example of stochastic simulation would be to sample many times from a complex space with many local maxima in order to find the global maximum. The estimate of the global maximum is then the maximum value of the sample generated by a long stochastic simulation. Samples from the complex space in question need not be independent as long as they visit the space proportional to its

support. Markov chains are one way of generating this long stochastic simulation. A sequence of random variables is called a Markov chain if, given the current state of the chain, the next state does not depend further on the history of the chain. Under certain conditions, as the number of samples increases, the chain will eventually converge to a stationary distribution from any initial value. [190] For a parameter, then, the mean of the stationary distribution is the expectation of the mean value of the parameter.

There are different sampling algorithms under the umbrella of MCMC, all designed to provide a chain that converges to a desired target distribution which is typically the posterior distribution of the parameters of interest. The Metropolis-Hastings algorithm works as follows. If the current position of the sampler is X, and the conditional density around this point $\pi(X)$, we sample a proposal Y from a distribution $q(Y|X)$ which is symmetrical around X. We accept the proposal Y with probability

$$\alpha(X,Y) = min\left(1, \frac{\pi(Y)q(X|Y)}{\pi(X)q(Y|X)}\right)$$

See [191]. If Y is rejected, the chain does not move. Gibbs sampling is a special case of Metropolis-Hastings sampling that takes as the proposal distribution the full conditional of a parameter or set of parameters. [192] Gibbs sampling therefore always accepts the proposal.

3.6. Informative priors

In a Bayesian analysis, all parameters have priors. Non-informative priors (e.g. a Gaussian distribution with a very large variance) are used in an attempt to approximate no prior information as to the value of the parameter. In simple situations the use of non-informative priors approximates a frequentist analysis. Non-informative priors can be problematic since they assign approximately equal prior probability both to realistic values of the parameter and to unrealistic values; if there is little information in the data the sampler may have difficulty visiting the resulting diffuse posterior distribution. Prior information independent of the data being analyzed can be used to restrict values of particular parameters to a plausible range (so called weakly informative priors) or to place higher prior probability on some values and very low prior probability on others (strongly informative

priors). Informative priors can help two chains for a parameter mix better if they were having difficulty mixing under the diffuse reference prior. Little information in the data will mean posteriors are dominated by priors; sensitivity of posteriors to choice of prior should be checked.

3.6.1. Prior setting

There is a vast literature on the topic of eliciting expert opinion for setting priors in a Bayesian analysis (e.g. [193–196]). Focus group discussions may be held and individuals asked to define median and 95% cut-points on the value of a parameter of interest. This process requires participants to be trained in how to express their beliefs in the form of a probability distribution. Alternative methods, where participants are asked to indicate points within a range where they are ambivalent as to whether their belief lies to the left or right of the line can be useful for establishing a distribution around a person's belief. These processes are labour-intensive and must be budgeted for at the planning stages of a study.

As an alternative to eliciting expert opinion, priors can be set based on a literature search. In the thesis, weakly informative priors for the effect of cluster transitions/antigenic distance between clusters, and for vaccine impact, on mean shifts were set using estimates of the variability in excess mortality and morbidity across influenza seasons. [7, 8, 54, 56, 58, 61, 65, 73, 92, 107, 197–199] A strongly informative prior on the effect of cluster transitions/antigenic distance on the mean shift was informed by a modeling study. [14] In all cases, data used to set priors were independent of data analyzed in the thesis.

3.7. Diagnosing convergence of Markov chain Monte Carlo

When the generated samples come from (approximately) the stationary distribution, the simulation is said to have converged. There are diagnostic plots that provide evidence against non-convergence to the target distribution (but they do not provide conclusive evidence for convergence). Henceforth evidence against non-convergence will be labeled apparent convergence. To show evidence of apparent convergence, two or more Markov

chains for a parameter can be run from different initial values and diagnostics used to determine if chains share a stationary distribution. Convergence of two chains of the LL can be monitored as a global indicator of convergence because the LL is contributed to by the parameters, the data and the priors. The LL of the Poisson HMM is

$$LL = \sum_{t=1}^{n} (-\mu_t + Y_t \log(\mu_t) - \log(Y_t!))$$

where μ_t is the predicted count, which is conditional on the state variable for that week, and Y_t is the observed count. One diagnostic plot for convergence is the Brooks-Gelman-Rubin (BGR) plot. [200] BGR plots monitor convergence of the ratio of variability of pooled chains to that within chains to 1. If chains have converged to the same stationary distribution, approximately all variability will be encompassed within the chains, with no additional variability between chains. BGR plots also show the width of the central 80% interval of the pooled runs and the average width of the 80% intervals within the individual runs, both of which should stabilise. For plotting purposes the pooled and within interval widths were normalised to have an overall maximum of 1 and statistics were calculated in bins of length 50. Figure 3.2 shows example BGR plots of apparent convergence and lack thereof. The ratio of pooled to within chains variability (top (red) line) is expected to be greater than 1 for early iterations of the sampler if the initial values for the two chains are suitably different from one another (i.e. the pooled variability far exceeds within chain variability).

For HMMs, key parameters are transition probabilities (δ) for the probability of moving between states ('aberrant' and 'normal') given the state of the previous week. Apparent convergence of transition probabilities typically predicts apparent convergence of the simulation more generally (e.g. figures 3.3 and 3.4).

3.8. Model adequacy

In this section several ways to demonstrate HMM model adequacy or superiority of one HMM over another are described. These are: posterior predictive density plots (section 3.8.1), the estimated time series of states ('normal' and 'aberrant') that gave rise to the observed data (section 3.8.2)

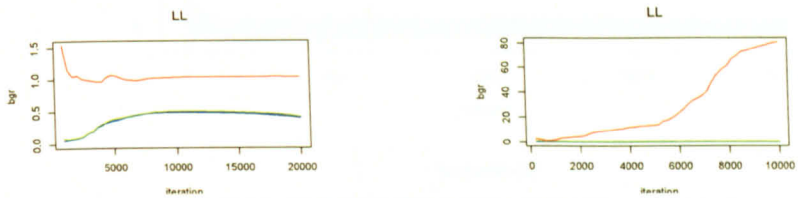


Figure 3.2.: **Examples of BGRs showing apparent convergence (left plot) and lack of convergence (right plot).** The red line is the ratio of pooled to within chains variability. If this line comes to 1, this is evidence for apparent convergence of the two chains to the target distribution. The two lines along the bottom of the plots are the width of the central 80% interval of the pooled runs (green) and the average width of the 80% intervals within individual runs (blue).

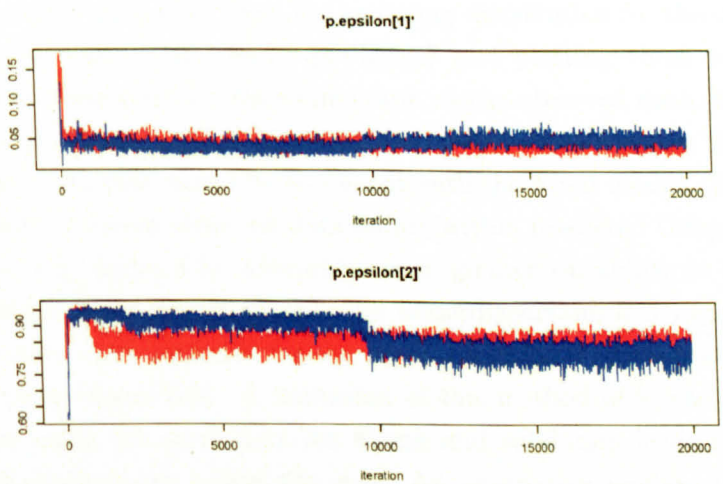


Figure 3.3.: **An example of a history plot of transition probability parameters showing apparent convergence to the same area of parameter space.**

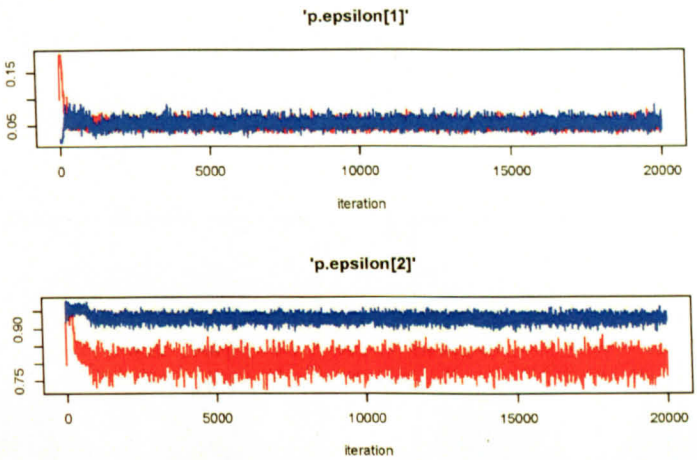


Figure 3.4.: **An example of a history plot of transition probability parameters showing lack of convergence.**

and autocorrelation plots of residuals (section 3.8.3).

3.8.1. Posterior predictive density plots

Posterior predictive density (PPD) plots were created by drawing a predicted count for each week from the sampling distribution for the observed data (whose mean is the linear predictor) and plotting these predicted counts, with their CrIs, on the same graph as the observed data. [201, 202] The CrI around predicted counts can be thought of as showing the range of observed data that would be consistent with the fitted model. [203] Adequate model fit (most observed data falling within predicted CrIs) is easily visualised (e.g. figure 3.5). Overdispersion, greater variability in observed data than predicted by the model, and underdispersion, less variability in observed data than predicted by the model, are also obvious from looking at PPD plots (figure 3.5). A limitation of this method of assessing model fit is that using the same data for fitting and validation leads to bias in favour of thinking the model fits well. An alternative method for model checking is to fit the model to a portion of the time series and predict the part not used in model fitting. [204] A limitation of this alternative approach for model checking is that it generally penalises complex models and models that make predictions too close to the sample data but unstable out-of-sample predictions. [205, 206]

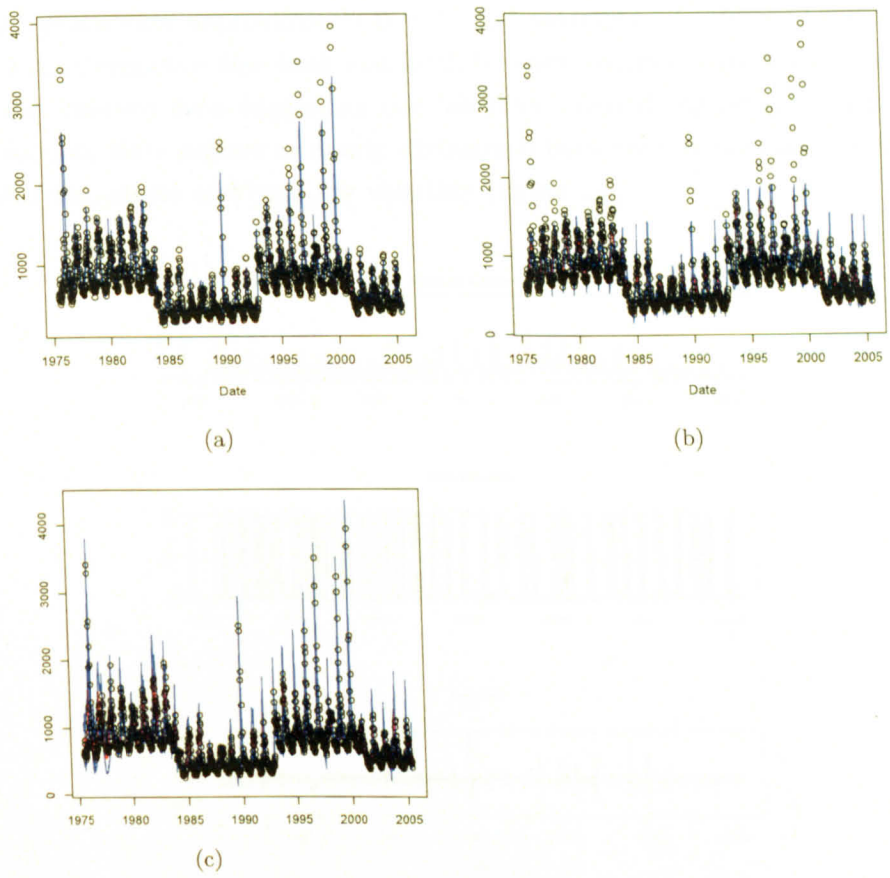


Figure 3.5.: **Posterior predictive density plots for (a) an adequate model fit and model fits showing (b) overdispersion (many observed data falling outside posterior predictive CrIs) and (c) underdispersion (no observed data outside the posterior predicted CrIs).** Circles: observed data, solid lines: median posterior predicted count for each week with 95% CrI.

3.8.2. State sequences

Models were compared with respect to their state sequences: the estimated time series of states ('normal' and 'aberrant') that gave rise to the observed P&I or ILI data (figures 3.6 to 3.8). State sequences were assessed in terms of apparent convergence (each week being assigned to the 'aberrant' state with probability approximately 0 or 1) and volatility (the degree to which the state sequence flips back and forth between states during a season despite observed data suggesting one 'aberrant' period during the season). Adequate state sequences clearly distinguish between the 'normal' and the 'aberrant' states and have low volatility (figure 3.6).

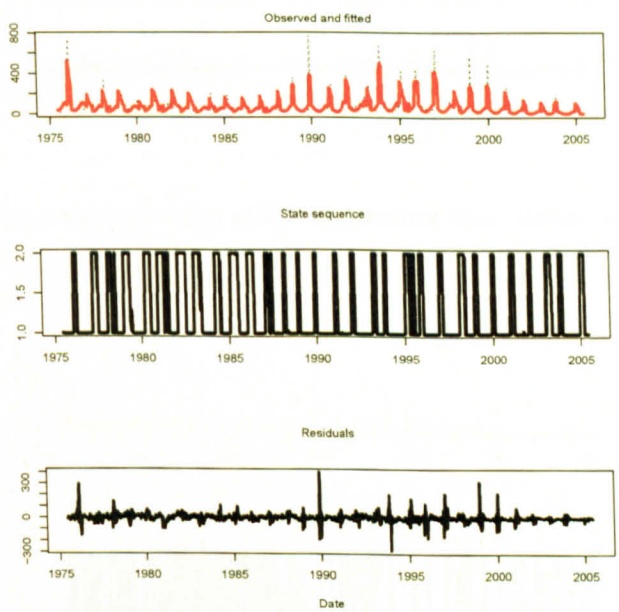


Figure 3.6.: **An example of a state sequence that appears to converge and is clearly estimated.** Top panel: observed and fitted data; middle panel: state sequence (1.0: 'normal' incidence, 2.0: 'aberrant' incidence); bottom panel: residuals (observed minus fitted count for each week).

3.8.3. Residual autocorrelation

An adequate model fit captures most of the variability in the observed data and leaves little serial correlation in residuals. Autocorrelation plots of

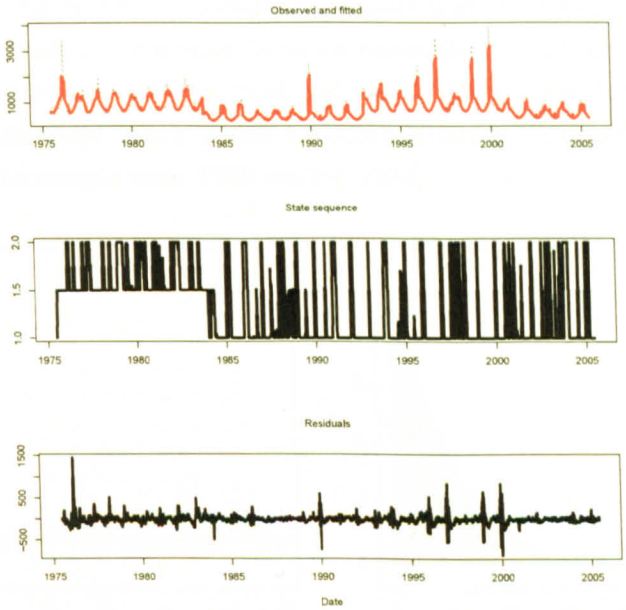


Figure 3.7.: An example of a state sequence that does not converge.

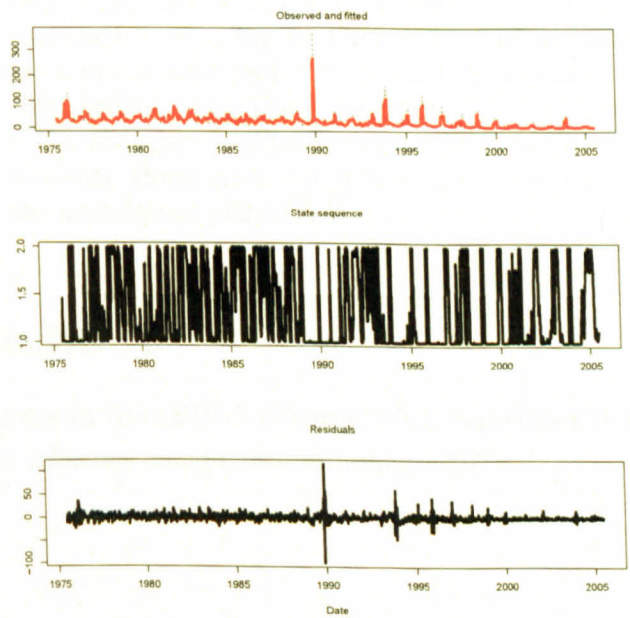


Figure 3.8.: An example of a state sequence that appears to converge but is volatile.

residuals separated by between 0 and 120 weeks showed the degree of positive correlation between residuals at short lags, which when present suggests underfitting of large peaks in the data, and at seasonal lags (approximately 52 and 104 weeks), which when present suggests inadequate fitting of seasonality (e.g. figure 3.9). Substantial correlation at other lags suggests poor model fit generally. Correlation between residuals is 1 at lag 0 because this correlation is between the residual and itself. Horizontal dotted lines indicate the threshold below which autocorrelation is 'ignorable' ($\pm 2/\sqrt{n}$, where n is the sample size: 1566 weeks). [204]

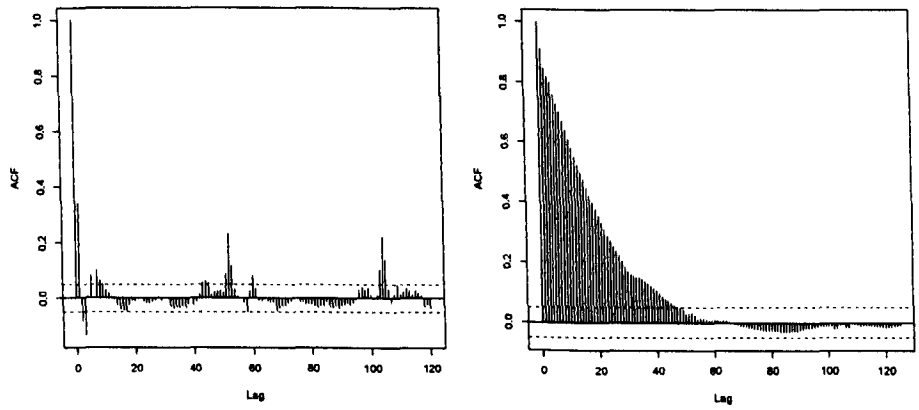


Figure 3.9.: **Example plots of the autocorrelation function of residuals. (a) An adequate model fit with some positive correlation at a lag of 1 week and at seasonal lags (approximately 52 and 104 weeks), but otherwise minimal correlation. (b) Poor model fit.** Y-axes: correlation between residuals at different lags, x-axes: lag between residuals, in weeks. Horizontal dotted lines are set at 'ignorable' residual autocorrelation ($\pm 2/\sqrt{1566}$).

3.9. OpenBUGS

MCMC was run in OpenBUGS (Version 3.0.2, September 2007), a package for Bayesian inference using Gibbs sampling. [207]

3.10. R

Data were first read into R (R version 2.6.2 (2008-02-08)) and OpenBUGS was called from within R to do the MCMC sampling using the BRugs package. Diagnostics and plots were produced in R.

4. Descriptive Results

4.1. Aims of this chapter

The aims of the work described in this chapter were, first, to determine the best way to model long-term trend in P&I and ILI incidence between 1970 and 2005 in England & Wales and, second, to explore crude associations between the peak incidence each influenza season and the exposures of interest: CT seasons, the antigenic distance between clusters and the vaccine coverage of those ≥ 65 each influenza season.

4.1.1. Objectives of this chapter

1. To fit negative binomial GLMs to P&I and ILI from 1970 to 2005 to determine whether long-term trend is adequately modeled by a linear term or if a quadratic term or cubic splines are needed

2. To determine if there is an interaction between age group and long-term trend in P&I or ILI

3. To explore crude associations between peak incidence each season and CT seasons, the antigenic distance between clusters and the vaccine coverage of those ≥ 65 each influenza season

4.1.2. Main findings

The long-term trends in rates of ILI and P&I over the past 36 years are not linear. Of the models tested, trend is best modeled by cubic splines with 5 df for P&I and 14 df for ILI. This is thus how long-term trend was fitted in HMMs in chapters 5 to 7. Long-term trends differ between P&I and ILI and across age groups for each outcome, necessitating age group-specific models in subsequent chapters. Up to 25% of the highest counts were excluded to isolate the long-term trend in 'normal' incidence from the influence of

'aberrant' observations. The shape of the long-term trend in both data sets is relatively insensitive to the exclusion of data. Overdispersion and autocorrelation are evident in residuals from models excluding even 25% of the highest counts.

In general, the distribution of peak rates for CT seasons appears greater than for intracluster seasons. Ranking influenza seasons in terms of the peak P&I or ILI rate, by age, shows that fewer than half of the top ten seasons are CT seasons. T-test results suggest weak evidence for, crudely, peak incidence observed in CT seasons being greater than that observed in the average season by approximately 6 deaths per 1,000,000 population ($p > 0.1$) and 95 consultations per 100,000 population ($p > 0.1$). These differences are not of public health relevance. There is no clear association between peak rates of P&I and ILI each influenza season and the antigenic distance between clusters. There is a weak negative association between peak P&I and ILI during an influenza season and vaccine coverage of the ≥ 65 age group in that season.

4.2. Introduction

In section 2.3.1 it was shown that influenza - attributable morbidity and mortality is variable season to season and that no consistent (increasing or decreasing) trend has been observed. [1, 53] In section 2.7, coincidence between large antigenic drift events and epidemic morbidity and mortality was noted. [58, 63, 93, 94, 107–110] A model of the genetic and antigenic evolution of influenza A/H3N2 coupled with a transmission dynamic model suggests that CT seasons may result in an average of 1.6 times higher peak influenza incidence than intracluster seasons. [14] The lack of an estimate of the average effect of CT seasons (as identified by Smith *et al.* [13]) was highlighted.

In section 2.8.3 it was noted that vaccine coverage of those ≥ 65 in England & Wales each influenza season increased from 24% in 1989/90 to 71% in 2004/05. [119,120] The population level impact of this high coverage has not been assessed in either the ≥ 65 age group or across age groups. In section 2.9, studies of the impact of rising vaccine coverage of the elderly on morbidity and mortality in the elderly were reviewed. After adjusting for confounding, population impact of high vaccine coverage in the elderly

is hard to detect by linear regression of excess mortality estimates. [7, 67] Results presented by Jansen *et al.* and Kwong *et al.* would suggest that for an increase in coverage of the elderly of approximately 20-30%, a decline in influenza - attributable mortality of 35-70% could be expected. [64, 95] Despite careful control for confounding, it is unlikely that these estimated declines in mortality are entirely attributable to vaccination. [133]

4.3. Data sets

An overview of the data analyzed is shown in figure 4.1. Figures 4.2 and 4.3 show weekly rates of ILI per 100,000 and P&I per 1,000,000 by age group (0-4, 5-14, 15-44, 45-64 and $\geq$ 65) from 1970 to 2005. Vaccine coverage in those $\geq$ 65 from 1989/90 to 2004/05 was derived from published sources (see section 3.2.4 in chapter 3). [119, 120] Information on whether influenza seasons were dominated by influenza A/H3N2 virus was taken from the literature (table 3.2). Information on which seasons were CT seasons was taken from Smith *et al.* (table 3.2). [13] From these two pieces of information a list of the first H3N2-dominated seasons after a CT was derived . The antigenic distance between clusters was also taken from the Smith paper. [13]

4.4. GLM methods

Negative binomial models with a log-link were fitted separately to weekly counts of P&I and ILI by age group from 1970 to 2005. [181]

$$\begin{aligned} Y_t &\sim \text{Poisson}(\mu_t\theta) \\ \log(\mu_t) &= \alpha + \beta t \\ \theta &\sim \text{Gamma(a, b)} \end{aligned}$$

Y_t is the observed P&I or ILI count in week t, which is a Poisson random variable with mean $\mu_t\theta$. The parameters of the Gamma distribution that θ follows control overdispersion; values of θ greater than 1 indicate the data are overdispersed relative to the Poisson model. The $\log(\mu_t)$ was modeled as dependent on N_t, the population at risk in week t, [180] α_1, the intercept, $C(t, \varphi)$, a cubic spline with φ df to model trend, $\beta_1 \sin\frac{2\pi t}{52.2} + \beta_2 \cos\frac{2\pi t}{52.2}$, representing seasonality, [70] $\beta_3$artifacts, the effects of known artifacts and

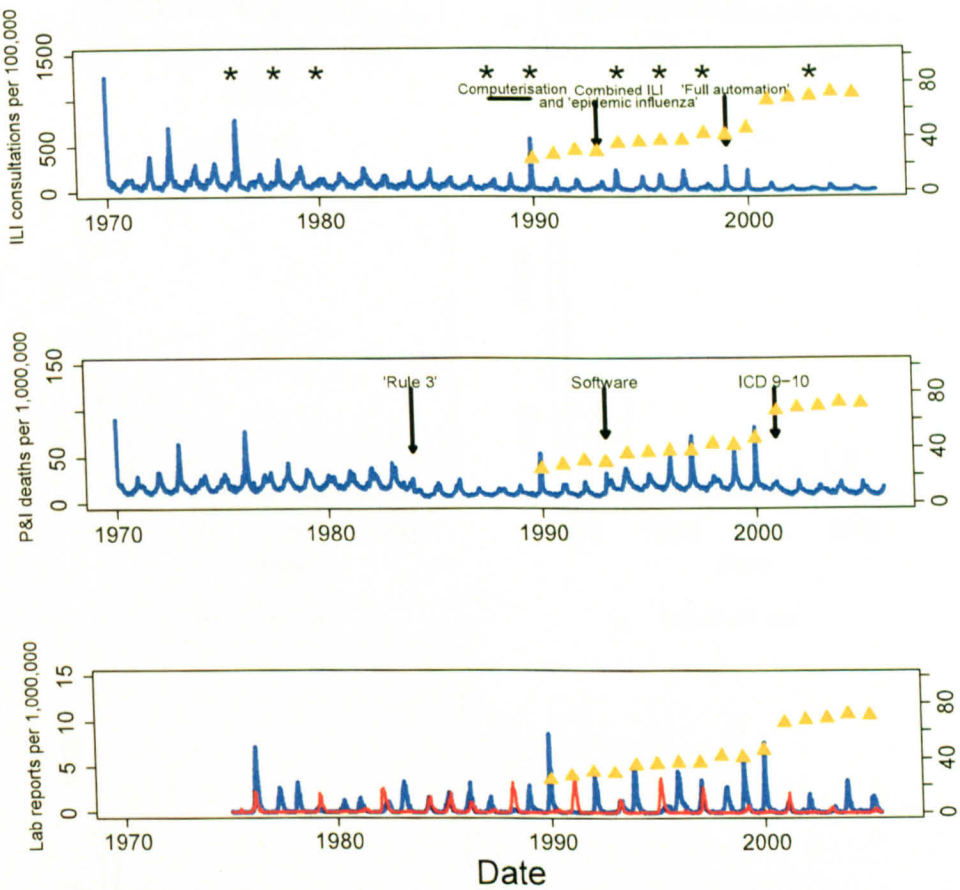
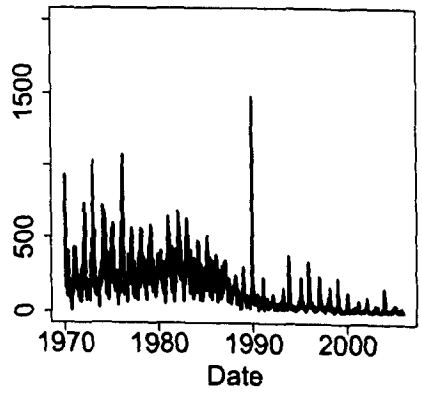
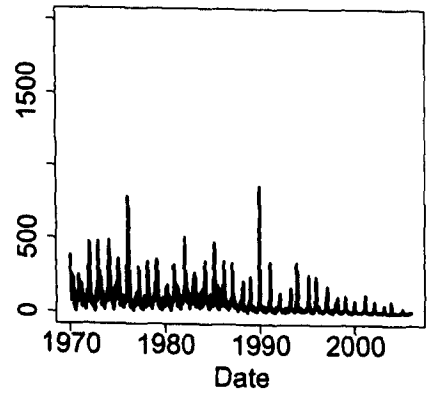


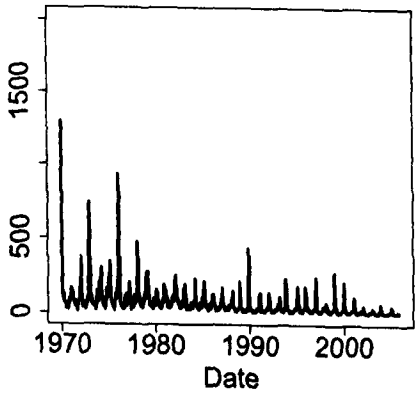
Figure 4.1.: **All-age weekly rates of ILI (top), P&I (middle) and laboratory reports for influenza A (bottom, blue) and B (bottom, red) in England & Wales from 1970 to 2005.** Stars indicate CT seasons. Triangles are vaccine coverage in the ≥ 65 (right y-axis). Arrows indicate artifacts.



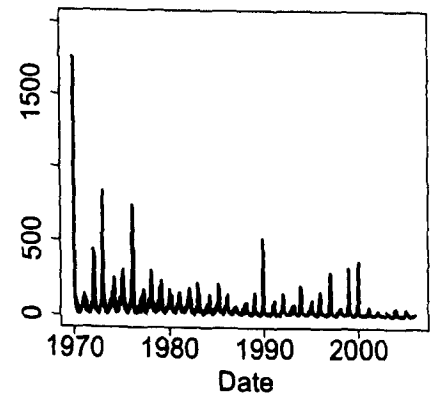
(a) 0-4 yrs



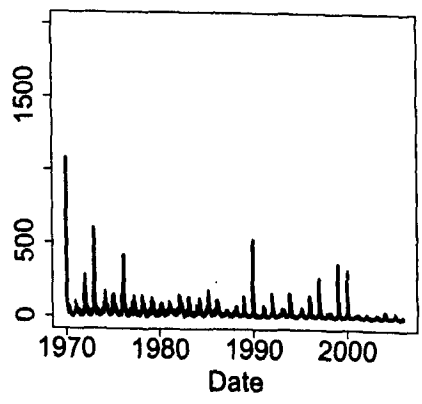
(b) 5-14 yrs



(c) 15-44 yrs

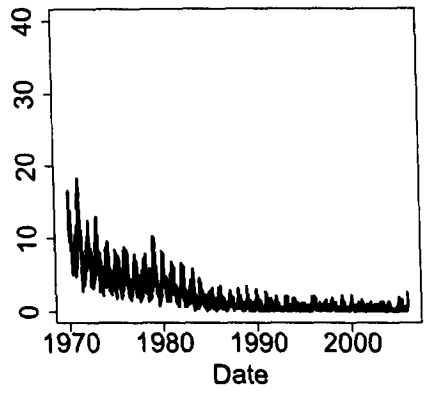


(d) 45-64 yrs

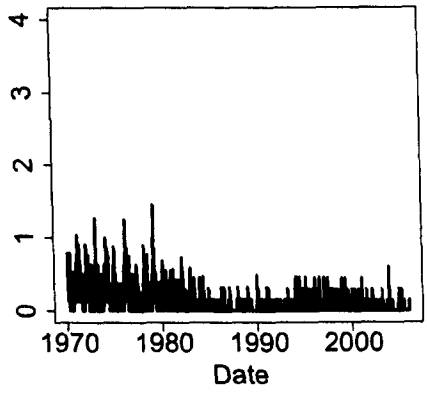


(e) $\geq$ 65 yrs

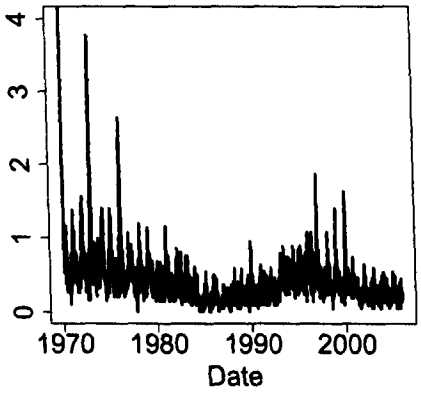
Figure 4.2.: **Weekly rates of GP consultation for ILI per 100,000 from the RCGP WRS, by age group: 1970-2005.**



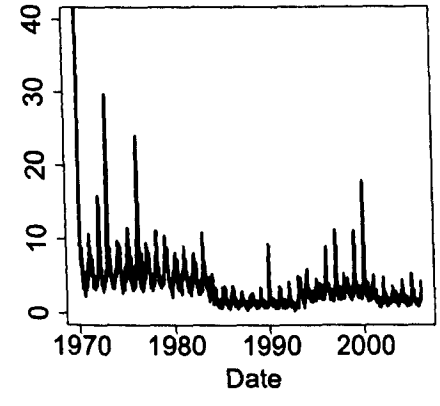
(a) 0-4 yrs



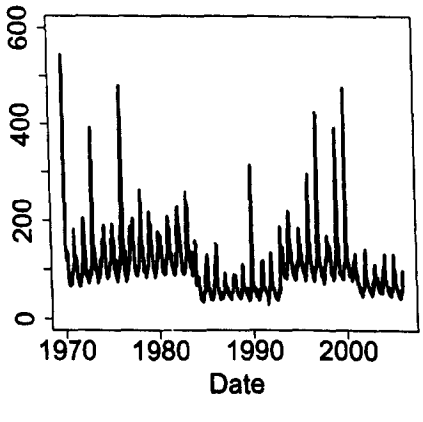
(b) 5-14 yrs



(c) 15-44 yrs



(d) 45-64 yrs



(e) $\geq$ 65 yrs

Figure 4.3.: **Weekly rates of P&I per 1,000,000 in England & Wales, by age group: 1970-2005.** Note different scales across age groups.

interaction terms between age group and artifacts (age group*artifacts) and between age group and trend (age group $* C(t, \varphi)$). The population at risk each week was the registered population of the sentinel general practices for ILI and the census population of England & Wales for P&I.

To fit the long-term trend in the data, extreme counts were removed. This is because different biological, environmental and behavioral factors drive large epidemics than drive long-term changes in incidence. Robustness of the fitted long-term trend to removal of 1, 2.5, 5, 10, 15, 20 and 25% of the highest counts in the two data sets was assessed.

Once high rates were excluded, long-term trend was modeled as linear or quadratic or flexibly with cubic splines with 4 to 20 df; more df were not tested to avoid overfitting the data. Interaction terms between age group and trend terms were included to test for a different long-term trend across age groups (0-4, 5-14, 15-44, 45-64, $\geq$ 65 years). Age group-specific models were also fitted to plot age group-specific trend lines.

AIC were plotted against the model to show thresholds at which additional model complexity did not lead to improved goodness of fit. AIC were also compared to test improvement of model fit with age-trend interaction included.

4.5. Descriptive methods

The peak weekly rate of P&I and ILI for each age group in each influenza season was plotted against vaccination coverage of the $\geq$ 65, the first H3N2-dominated season after a CT and the antigenic distance between CT seasons to visualise crude associations. One and two-sample t-tests and a permutation test (assuming each season is independent i.e. no secular trend in peak seasonal incidence) were performed to determine whether crude associations between CT seasons and peak incidence might be due to chance. [208] One-sample t-tests addressed whether the difference between mean peak incidence observed in CT seasons and mean peak incidence observed in all seasons combined is greater than 0. Two-sample t-tests were used to determine whether the difference between mean peak incidence in CT seasons and the mean peak incidence in non-CT seasons was different from 0. The permutation test was done to determine how many times the same difference between mean peak incidence in CT seasons and mean peak incidence

in all seasons, or greater, was observed if CTs were randomly allocated to 10 of the 36 seasons in the data set. One million replicate samples of 10 seasons were drawn and p-values calculated summarising the number of times the mean peak incidence in the sampled seasons was greater than the mean peak incidence across all seasons.

4.6. Long-term trend in P&I and ILI

The long-term trend in P&I and ILI refers to the secular trend in P&I and ILI not attributable to influenza. Recall that the trend in these data that is not attributable to influenza was isolated by excluding high counts, which were assumed to be influenza-attributable, from model fits. Model fit to the long-term trend in P&I and ILI improved with each additional df for the cubic splines up to 14 df for the ILI data and 5 df for the P&I data (figures 4.4 and 4.5). Allowing an interaction between the long-term trend and age group improved model fit for both ILI and P&I data.

Long-term trend in weekly ILI rates declined between the mid 1980s and mid 1990s, leveled off and declined again from approximately 2000 in all age groups (figure 4.6). In a supplementary analysis, similar declines since 2000 were observed for other upper and lower respiratory tract infections, as well as non-respiratory consultation categories (figures A.1 to A.8).

Dramatic declines in the long-term trend in weekly P&I rates in children were observed (figure 4.7). For the 15-44, 45-64 and $\geq$ 65 age groups, the 95% CIs around the fitted long-term trend curves are consistent with some decline in the long-term trend in P&I since approximately 1998 for those in the 15-44 and $\geq$ 65 age groups, and after 2000 for the 45-64 age group. Near the end of the time series, CIs for the 15-44 and 45-64 year age groups are wide and encompass the possibility of a rise or a fall in the long-term trend. The CI around the long-term trend for the oldest age group is consistent only with a decline in the long-term trend in P&I between 1998 and 2005.

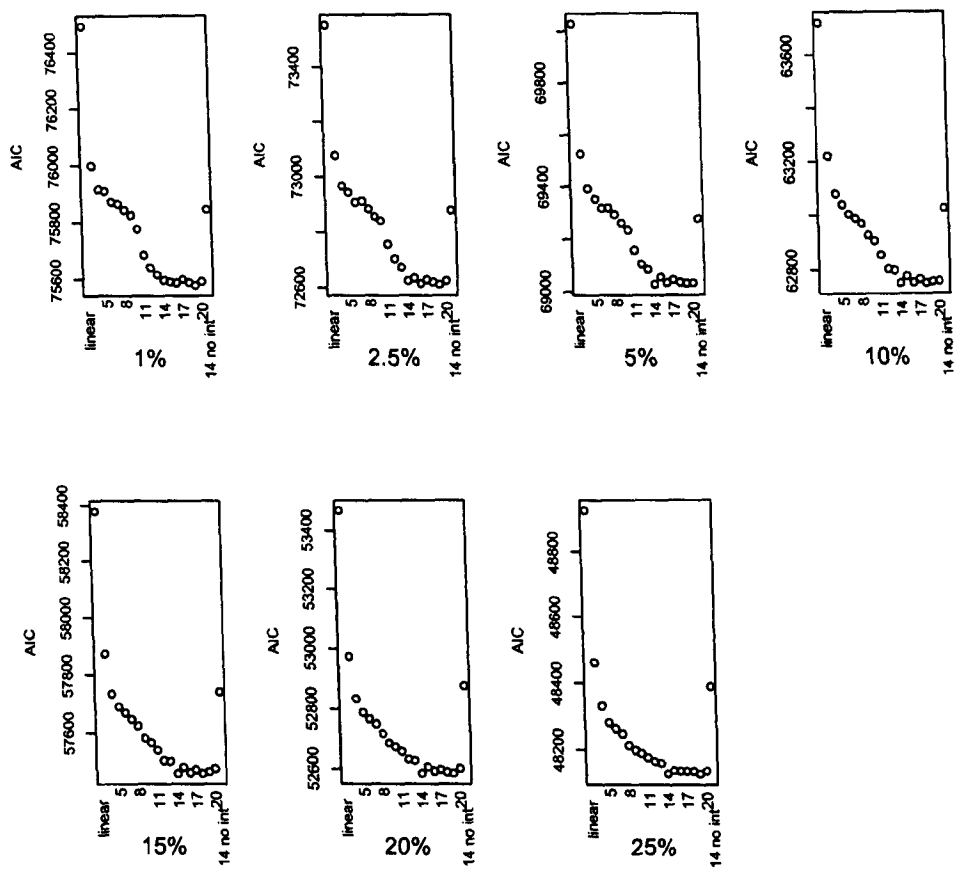


Figure 4.4.: **Model selection: negative binomial GLMs fitted to ILI.** AIC on the y-axis against model complexity (linear or quadratic trend or cubic splines with up to 20 df) on the x-axis. Each plot is a different percentage of high counts excluded from model fit (1-25%). AIC do not fall appreciably beyond a model where the long-term trend is modeled with cubic splines with 14 df. The last point on the far right of the plot is the AIC for the model excluding age group-trend interaction, where trend is modeled as cubic splines with 14 df.

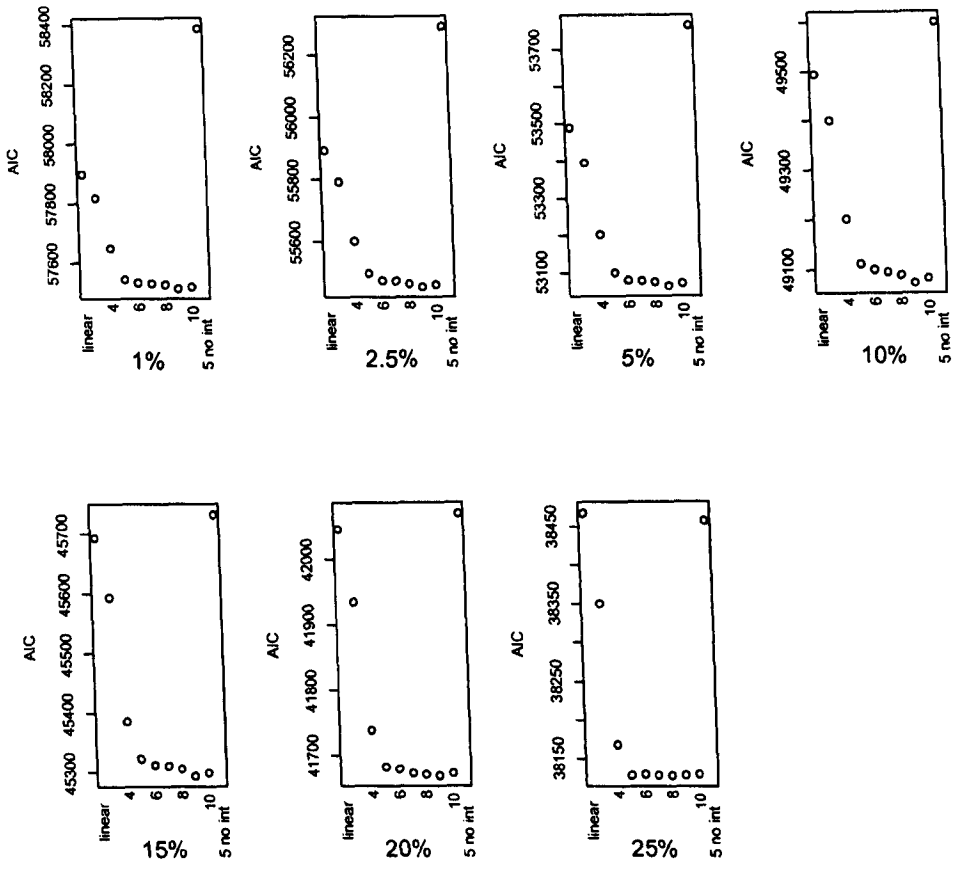
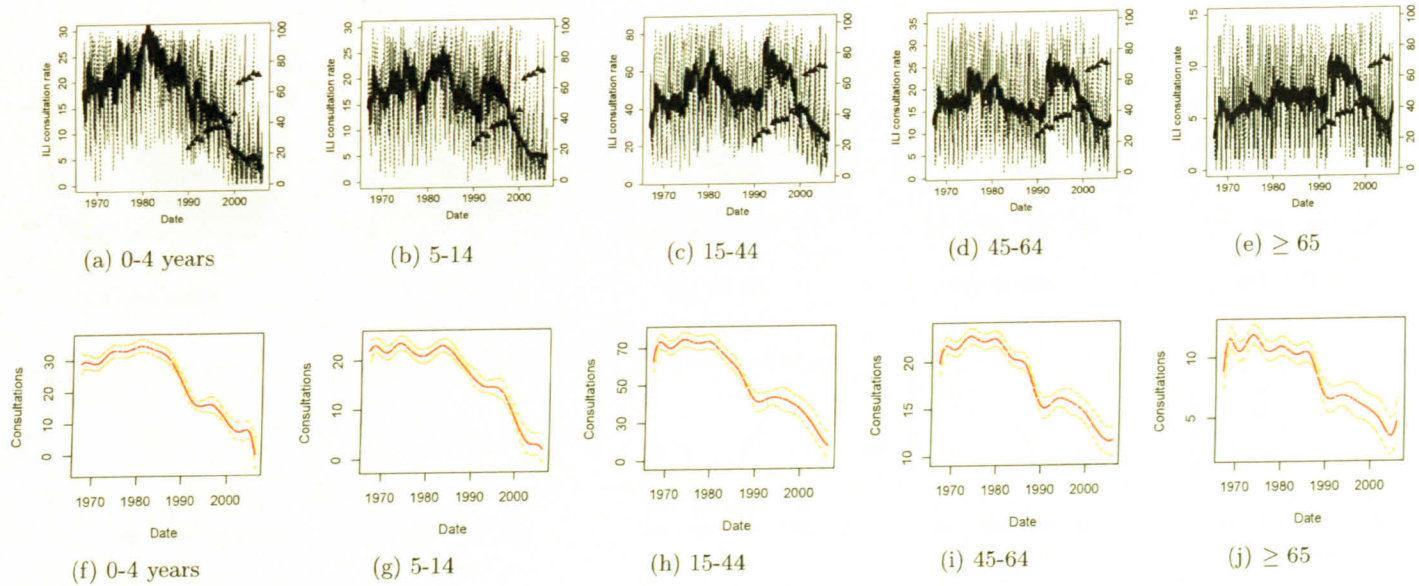


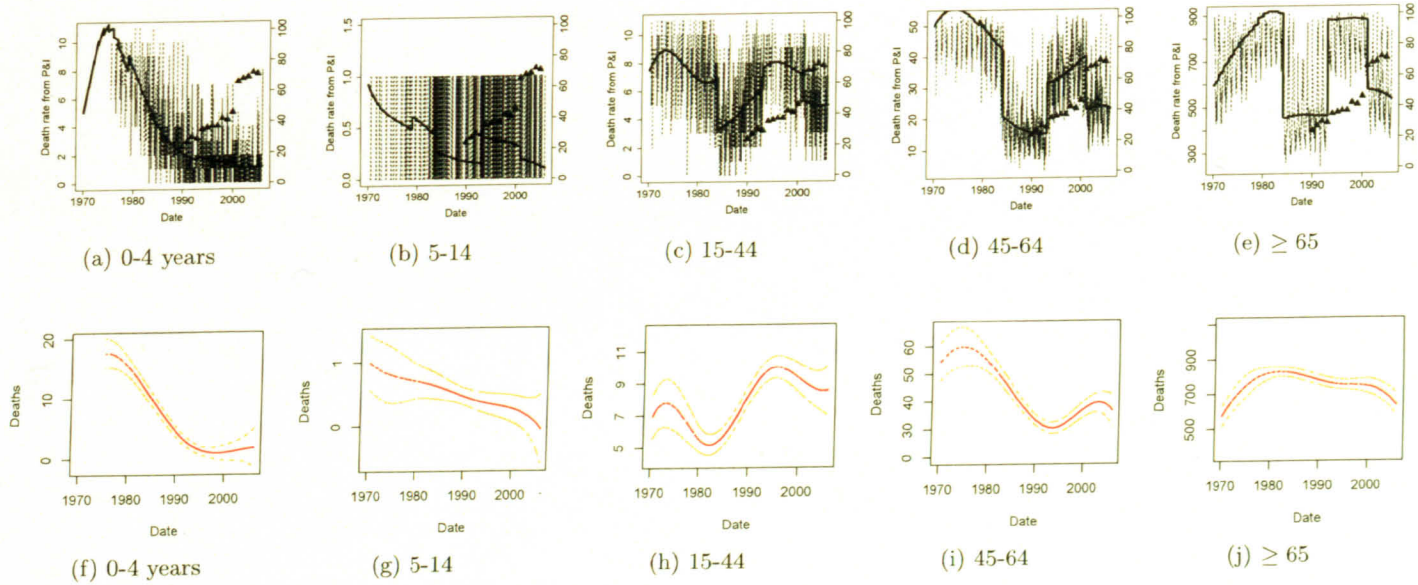
Figure 4.5.: **Model selection: negative binomial GLMs fitted to P&I.** AIC do not fall appreciably beyond a model where the long-term trend is modeled with cubic splines with 5 df. The last point on the far right of the plot is the AIC for the model excluding age group-trend interaction, where trend is modeled as cubic splines with 5 df.



(a) 0-4 years (b) 5-14 (c) 15-44 (d) 45-64 (e) ≥ 65

(f) 0-4 years (g) 5-14 (h) 15-44 (i) 45-64 (j) ≥ 65

Figure 4.6.: **Top panel: The fitted long-term trend in ILI by age group. Bottom panel: The cubic spline component of the fitted long-term trend in ILI (red curve) and a 95% CI about the fitted spline (gold curves).** 25% of the highest counts were excluded from model fitting. In the top panels, visible step changes in the long-term trend indicate known artifacts in the data set that were allowed for in estimating the shape of the long-term trend. Triangles indicate influenza vaccine coverage in those aged $\geq$ 65 (right y-axis). [120, 209]

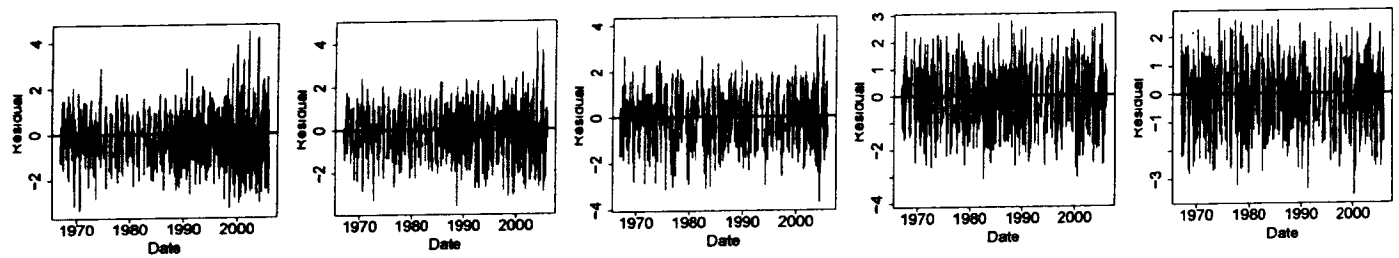


(a) 0-4 years (b) 5-14 (c) 15-44 (d) 45-64 (e) ≥ 65

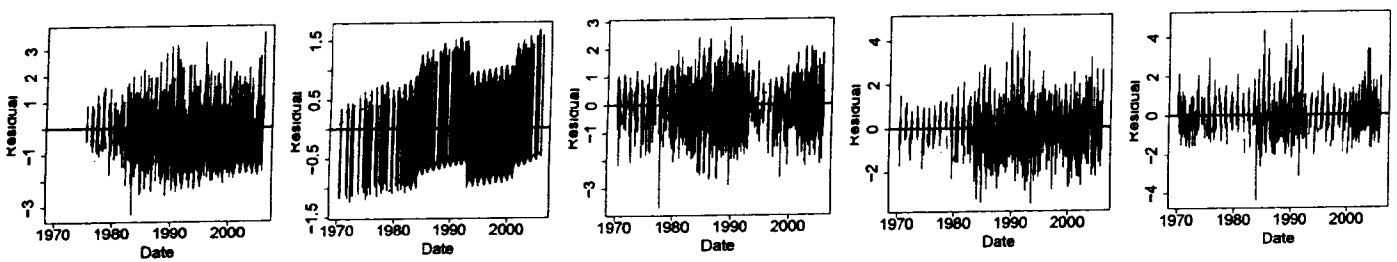
(f) 0-4 years (g) 5-14 (h) 15-44 (i) 45-64 (j) ≥ 65

Figure 4.7.: **Top panel: The fitted long-term trend in P&I by age group. Bottom panel: The cubic spline component of the fitted long-term trend in P&I (red) and a 95% CI (gold).** 25% of the highest counts were excluded from model fitting. For P&I in the 0-4 age group, exclusion of 25% of the highest counts resulted in all weeks of the years 1970-1975 being excluded from model fit.

Residuals from the models fitted to both data sets with 25% of the highest counts excluded are approximately Normally distributed around zero with some autocorrelation (figure 4.8); large residuals are evident from the fit to P&I data. Overdispersion was reduced with the exclusion of high counts, but large residuals, and values of θ greater than 0 for most models (table 4.1), with the exclusion of up to 25% of counts suggest it was not eliminated.



(a) ILI 0-4 (b) ILI 5-14 (c) ILI 15-44 (d) ILI 45-64 (e) ILI ≥ 65



(f) P&I 0-4 (g) P&I 5-14 (h) P&I 15-44 (i) P&I 45-64 (j) P&I ≥ 65

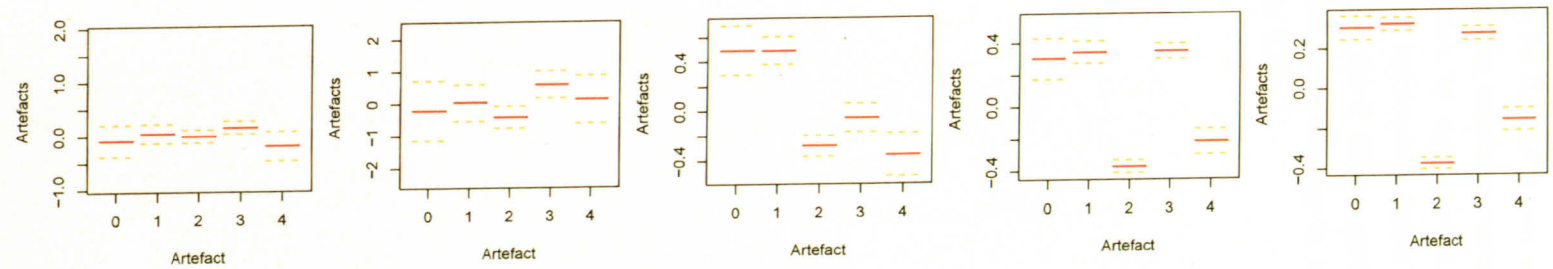
Figure 4.8.: **Residuals from negative binomial GLM fits to ILI and P&I excluding 25% of the highest counts from model fitting.**

Table 4.1.: **Values of dispersion parameter, θ, from final GLM models.** Blanks in the table mean that θ was not estimable from this model.

Outcome	spline df	age group	θ	se(θ)
P&I	5	0-4		
P&I	5	5-14		
P&I	5	15-44		
P&I	5	45-64	80.500	12.100
P&I	5	≥ 65	83.070	3.560
ILI	14	0-4	11.056	0.788
ILI	14	5-14	8.618	0.559
ILI	14	15-44	15.265	0.772
ILI	14	45-64	16.270	1.250
ILI	14	≥ 65	18.360	2.780

The shape of the long-term trend in P&I and ILI rates is relatively insensitive to the percentage of high rates excluded (figures B.1 and B.2).

Effects of artifacts differ across age groups (figure 4.9). There is little information in the data as to the impact of most of the artifacts (as evidenced by the overlapping CIs for most of them) except for P&I in the age groups 15-44, 45-64 and ≥ 65; the impact of these changes has been discussed by ONS. [154]



(a) 0-4 years (b) 5-14 (c) 15-44 (d) 45-64 (e) ≥ 65

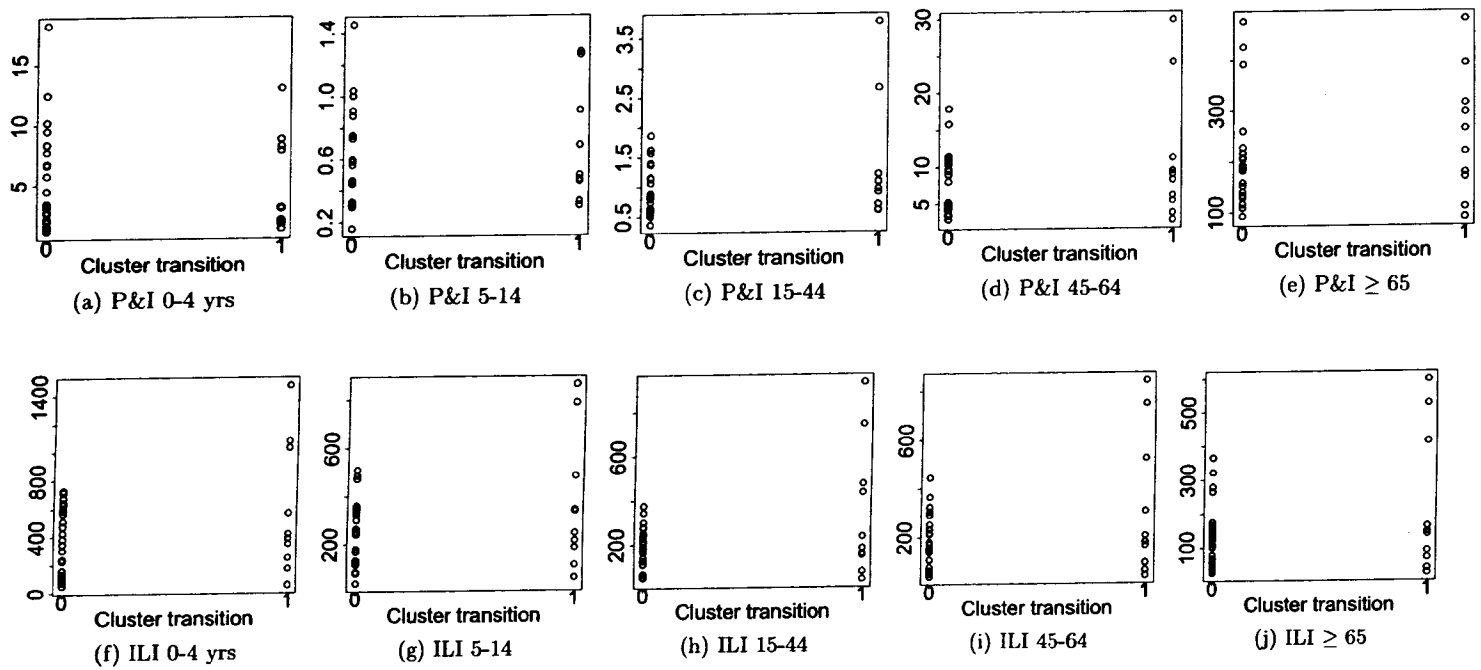
(f) 0-4 years (g) 5-14 (h) 15-44 (i) 45-64 (j) ≥ 65

Figure 4.9.: **Relative impact of known artifacts in P&I ((a)-(e)) and ILI ((f)-(j)) on the long-term trend, by age group from model fits excluding 25% of the highest counts.** The red line is the estimate of impact and the golden dotted lines above and below the solid lines are 95% CIs about each estimate. Artifacts are numbered in chronological order as in table 3.1. Artifact 0 is the referent.

4.6.1. Effect of CT seasons on epidemics

In general, the distribution of peak rates for CT seasons appears greater than for intracluster seasons (figure 4.10). This is consistent across age groups for ILI but is only seen for P&I from the 15-44 and 45-64 age groups.

Ranking influenza seasons from largest to smallest in terms of the peak P&I rate, by age, showed that CT seasons do not feature prominently in the top ten seasons (table 4.2). Five of the top ten seasons in terms of P&I rate for the $\geq$ 65 age group were CT seasons. Fewer than half of the top ten influenza seasons in terms of peak ILI rate, by age, were CT seasons (table 4.3).



(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64 (e) P&I $\geq$ 65

(f) ILI 0-4 yrs (g) ILI 5-14 (h) ILI 15-44 (i) ILI 45-64 (j) ILI $\geq$ 65

Figure 4.10.: **Peak weekly P&I rate per 1,000,000 population (top plots) and ILI rate per 100,000 population (bottom plots) observed during each influenza season vs. whether the season was a CT season (1) or an intracluster season (0).**

Table 4.2.: **Top ten influenza seasons in terms of peak P&I rate.** 1 is a CT season, 0 an intracluster season.

	0-4 yrs		**5-14**		**15-44**		**45-64**		**≥ 65**	
rank	**season**	**CT?**	**season**	**CT?**	**season**	**CT?**	**season**	**CT?**	**season**	**CT?**
1	1970	0	1978	0	1972	1	1972	1	1975	1
2	1972	1	1972	1	1975	1	1975	1	1999	0
3	1971	0	1975	1	1996	0	1999	0	1996	0
4	1978	0	1970	0	1999	0	1971	0	1972	1
5	1973	0	1973	0	1971	0	1974	0	1998	0
6	1975	1	1971	0	1998	0	1996	0	1989	1
7	1974	0	1977	1	1973	0	1977	1	1995	1
8	1979	1	1974	0	1974	0	1998	0	1977	1
9	1977	1	1976	0	1970	0	1982	0	1982	0
10	1976	0	1981	0	1977	1	1970	0	1981	0

Table 4.3.: **Top ten influenza seasons in terms of peak ILI rate.**

1	1989	1	1989	1	1975	1	1972	1	1972	1
2	1975	1	1975	1	1972	1	1975	1	1989	1
3	1972	1	1981	0	1977	1	1989	1	1975	1
4	1971	0	1973	0	1989	1	1971	0	1998	0
5	1973	0	1972	1	1971	0	1999	0	1999	0
6	1981	0	1984	0	1974	0	1998	0	1971	0
7	1980	0	1971	0	1973	0	1974	0	1996	0
8	1982	0	1974	0	1978	0	1977	1	1984	0
9	1974	0	1978	0	1998	0	1996	0	1973	0
10	1978	0	1983	0	1981	0	1973	0	1993	1

Crudely, peak incidence observed in CT seasons is greater than that observed in the average season by approximately 6 deaths per 1,000,000 population and 95 consultations per 100,000 population; t-tests suggest there is weak evidence that these are true differences (table 4.4). Peak incidence observed in CT seasons is greater than that observed in intracluster (non-CT) seasons by approximately 8 deaths per 1,000,000 population and 133 consultations per 100,000 population; these observations are less likely to be due to chance (table 4.4). In approximately 6% of permutations for the comparison of P&I peak incidence in CT seasons to the average season, at least as great a difference between CT seasons and the average season is observed as for the real CT seasons ($P = 0.062$). The peak weekly P&I rate in the data analyzed in the thesis, excluding the 1969/70 pandemic season, is 80.4 per 1,000,000 (week 52, 1999) and the peak weekly rate of ILI consultations is around 2,322 per 100,000 (week 49, 1989). Differences in the weekly peak of on average 8 P&I deaths per 1,000,000 population and 133 ILI consultations per 100,000 population between CT seasons and intracluster seasons are therefore not differences of public health importance.

Table 4.4.: **T-test and permutation test results.**

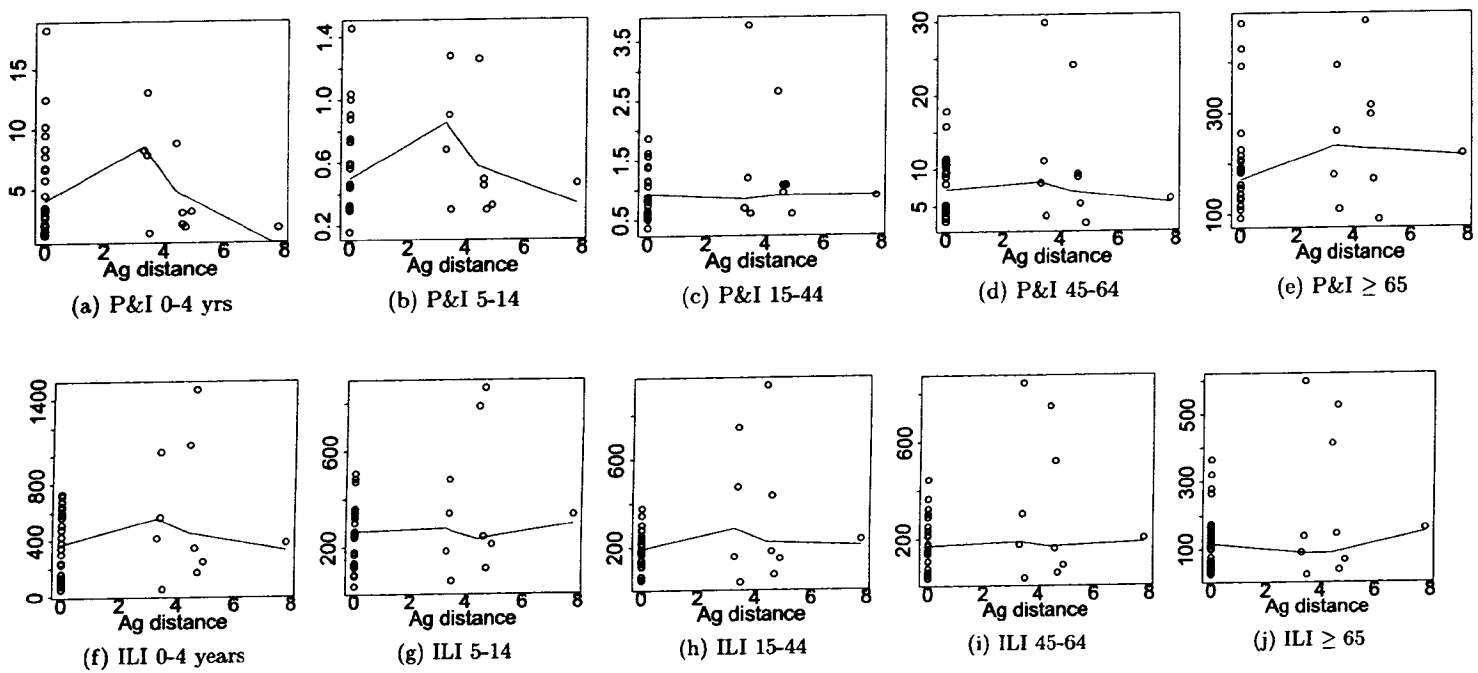
Test	t	df	P
One-sample t-tests			
P&I			
CT-seasons mean peak rate minus overall mean peak rate ≥ 0	0.27	9	>0.168
ILI			
CT-seasons mean peak rate minus overall mean peak rate ≥ 0	0.35	9	>0.168
Two-sample t-tests			
P&I			
CT-seasons mean peak rate minus non-CT season mean peak rate	1.65	33	0.12
ILI			
CT-seasons mean peak rate minus non-CT season mean peak rate	2.20	33	0.04
Permutation tests			
P&I			
CT-season mean peak rate vs. overall mean peak rate			0.062

Continued on Next Page...

Table 4.4 – Continued

Test	t	df	P
ILI			
CT-season mean peak rate vs. overall mean peak rate			0.024

There is no clear association between peak rates of P&I and ILI each influenza season and the antigenic distance between clusters (figure 4.11).



(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64 (e) P&I ≥ 65

(f) ILI 0-4 years (g) ILI 5-14 (h) ILI 15-44 (i) ILI 45-64 (j) ILI ≥ 65

Figure 4.11.: **Peak P&I rate (per 1,000,000) and peak ILI rate (per 100,000) in each influenza season vs. the antigenic distance between clusters (0 for intracluster seasons).** A lowess smoother was overlayed to visualise the associations.

4.6.2. Impact of vaccination on epidemics

There is a weak negative association between peak P&I and ILI during an influenza season and vaccine coverage of the ≥ 65 age group in that season (figure 4.12). The association is found across all age groups and appears stronger for ILI than for P&I.

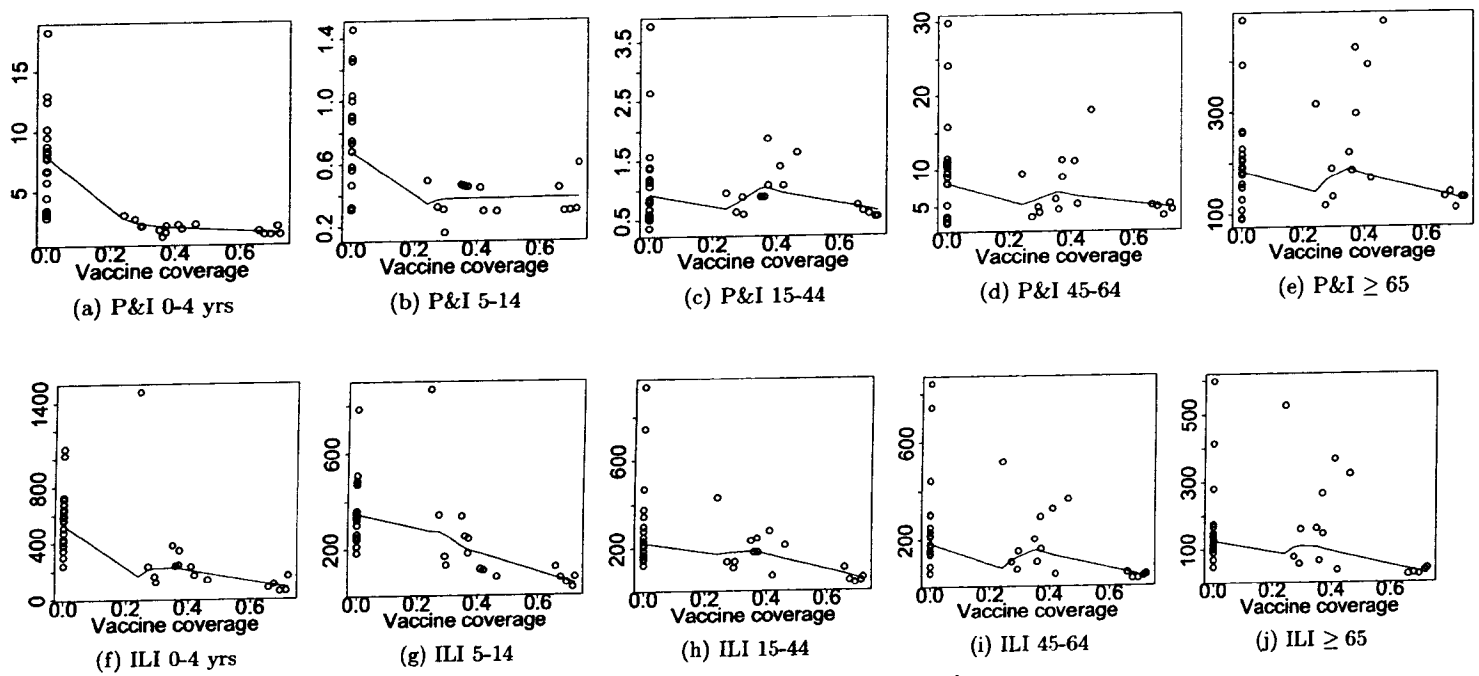


Figure 4.12.: **Peak P&I rate (per 1,000,000) and peak ILI rate (per 100,000) in each influenza season vs. the vaccine coverage of the ≥ 65 each influenza season.**

4.7. Summary of results

Long-term trend in rates of ILI and P&I between 1970 and 2005 are not linear. Of the models tested, trend is best modeled by cubic splines with 5 df for P&I and 14 df for ILI. Long-term trends differ between P&I and ILI. There is evidence for age group-trend interaction for both P&I and ILI. The data are consistent with a decline in weekly P&I rates in those $\geq$ 65 years old from approximately 1998 coincident with markedly increased yearly vaccination coverage in this age group. ILI consultation rates declined in all age groups between the mid 1980s and mid 1990s, were stagnant to 2000, and declined thereafter. The shape of the long-term trend in both data sets is relatively insensitive to the exclusion of data.

Crudely, the distribution of peak rates for CT seasons appears greater than for intracluster seasons. When peak P&I and ILI rates per season are ranked, only for P&I in the $\geq$ 65 age group do CT seasons occur in at least five of the top ten seasons. T-tests and permutation tests suggest that the small increases in peak P&I and ILI in CT seasons compared with the average season or intracluster seasons are supported by weak evidence and are of little public health importance. There is no clear association between peak rates of P&I and ILI each influenza season and the antigenic distance between clusters. There is a weak negative association between vaccine coverage of the $\geq$ 65 age group and peak seasonal P&I and ILI rates across age groups.

Peak rates of morbidity and mortality are sensitive to the baseline rate of morbidity and mortality (i.e. the rate of morbidity and mortality that would be expected in the absence of influenza). All things being equal (e.g. virulence of circulating influenza viruses, degree of vaccine match to circulating viruses, vaccine coverage, winter temperatures), if the baseline mortality rate declined, so would the peak mortality rate. Estimates of influenza - attributable morbidity or mortality, excess mortality for example, are less sensitive to baseline morbidity and mortality rates. In chapter 7, estimated mean shifts for P&I and ILI by age group for each season (the estimated influenza - attributable rate ratio of the 'aberrant' to 'normal' morbidity and mortality rates) from the HMMs are plotted against the same three metrics - CT vs. intracluster seasons, the antigenic distance between clusters and vaccine coverage in the $\geq$ 65. This gives a more

accurate estimate of crude association between the exposures of interest and influenza - attributable morbidity and mortality. Also in chapter 7, confounding of these associations of interest by other time-varying factors is discussed.

4.8. Strengths of the GLM methods used

Many factors may lead to an apparent change in the long-term trend in ILI or P&I. An example of such a factor is a secular change in use of GP services. Different factors than those explaining secular trends are responsible for variability of peak P&I or ILI incidence across influenza seasons. Examples of factors that may explain variability of peak P&I or ILI across influenza seasons are increased or decreased influenza vaccination and CTs. High rates at the start of the time series, for example in the years shortly following the first wave of the Hong Kong pandemic of 1968 when the modern influenza A/H3N2 viruses began to circulate in the human population, influence the slope of the estimated long-term trend (especially when fitting a linear trend). By excluding these high rates from model fitting and fitting a flexible curve to the remaining data using cubic splines a close approximation to the true shape of the long-term trend in P&I and ILI in England & Wales between 1970 and 2005 has been estimated.

4.9. Limitations of the GLM methods used

Excluding 25% of counts is insufficient to eliminate overdispersion and isolate the long-term trend in 'normal' incidence from 'aberrant' counts in the death data especially. Hidden Markov models may alleviate these problems by drawing 'normal' rates, that contribute to the long-term trend, from one distribution and 'aberrant' rates from a second distribution. [188]

4.10. Long-term trend

The decline in ILI since around 2000 in all age groups, a decline in other respiratory and non-respiratory consultation rates since 2000 and the absence of a similar decline in P&I rates across age groups indicate a behavioural

change has resulted in a fall in the use of GP services. Different factors probably contributed to the decline in use of GP services in different age groups. The reduction in provision of out-of-hours care by GPs in favour of deputies (whose notes are not incorporated into patient GP files), the introduction of the 'NHS Direct' telephone health line [210] and the *Path of least resistance* Department of Health report in September 1998 which discouraged antimicrobial prescribing in general practice [179] may have contributed to a decline in consulting by the young and working aged adults. [122] Policy in the early 1990s to redirect funding from residential care homes to supporting older people in their own homes probably resulted in elderly people also being less likely to consult their GP and more likely to be admitted to hospital if they developed a respiratory infection. [122,211] Progressively lower odds of death within 30 days of being admitted to hospital for pneumonia from 2000/01 to 2004/05 compared with 1997/98 reported by Trotter *et al.* supports this hypothesis of a lower threshold for admission of elderly people. [212] Trends in other respiratory disease consultations were similar which reinforces the suggestion that an environmental or a behavioural change may explain declines in ILI.

Reasons for the decline in mortality in the youngest age groups have been described elsewhere. [213] If all influenza-attributable P&I and ILI was not removed by deletion of 25% of the highest counts before model fitting, it is possible that the hint of a decline in P&I in the $\geq$ 65 age group after 1998 may be due to the increase in vaccine coverage in this age group. [120]

5. Univariate HMM results

5.1. Aims of this chapter

The aim of the work described in this chapter was to determine the appropriate scale of analysis and distributional assumptions for HMMs fitted to P&I and ILI and to evaluate what HMMs fitted to either P&I or ILI alone tell us about influenza.

5.1.1. Objectives of this chapter

1. To determine appropriate scale of analysis (log-link or identity link) and distributional assumptions (Poisson or negative binomial) for HMMs by first fitting each model to P&I for $\geq$ 65 age group and ILI for 15-44 age group and exploring convergence and model fit

2. To explore convergence and fit of good models from objective 1 to P&I for each age group and to ILI for each age group

3. Based on model fit and convergence from objective 2, to evaluate what these univariate models tell us about influenza

5.1.2. Main findings

Of the models fitted, Poisson log-link and identity link models were developed and other models discarded due to problems of convergence and model fit. Of the Poisson log-link and identity link models fitted to P&I and ILI for different age groups, state sequences (the model-estimated state assignment of each week in the time series) appeared to converge for Poisson identity link models fitted to P&I but not ILI. Poisson log-link model state sequences appeared to converge for fits to ILI but not P&I. P&I for all age groups apart from the $\geq$ 65 age group were underdispersed relative to both log-link and identity link models. ILI data were not underdispersed

or overdispersed relative to either log-link or identity link models. Evidence of apparent convergence of state sequences predicted apparent convergence of the model more generally (as captured by the LL). Both log-link and identity link models were taken forward for joint modeling with laboratory data (chapter 6).

At the end of the chapter, the concept of lack of convergence reflecting a lack of information in the data about state transitions is introduced. This idea is revisited in chapter 6 where multivariate models are built in an attempt to share information about state sequences, and therefore mean shifts, across outcome variables. It is revisited in chapter 7 when priors for the impact of CT seasons and vaccination of varying degrees of 'informativeness' are set. The amount of information in the data about the effects of interest determines the ease with which the sampler visits the different posteriors. Values of coefficients (e.g. for the magnitude of mean shifts) are not be interpreted in this chapter.

5.2. Introduction

In chapter 4, evidence for a nonlinear long-term trend in incidence of P&I and ILI was presented. Since the long-term trends in rates of ILI and P&I over the past 36 years are not linear, cubic splines were used to fit long-term trend in HMM model fits. Long-term trends differ between the two outcome variables, and across age groups for each outcome, necessitating age group-specific HMMs. Overdispersion and autocorrelation are not accounted for by fitting negative binomial GLMs to P&I and ILI excluding even 25% of the highest counts.

Peak ILI and P&I incidence is variable whether or not the season is dominated by a new cluster. Ranking influenza seasons in terms of the peak P&I or ILI rate, by age, showed that fewer than half of the top ten seasons were CT seasons. T-tests suggest weak evidence that peak incidence observed in CT seasons is greater than peak incidence observed in the average season by approximately 6 deaths per 1,000,000 population and 95 consultations per 100,000 population; a permutation test suggested that the observed difference between peak P&I incidence in CT seasons compared to the average season may be due to chance. There is no clear association between peak rates of P&I and ILI each influenza season and the antigenic

distance between clusters. There is a crude negative association between increasing vaccination coverage and seasonal peak ILI and P&I incidence observed across all age groups.

In chapter 3 the history of regression modeling of influenza was reviewed and the need to distinguish 'aberrant' from 'normal' incidence in time series' of non-laboratory confirmed morbidity and mortality was highlighted. This is usually done by excluding some portion of the data (sometimes informed by incidence of laboratory-confirmed influenza or by using cut-offs whereby a certain percentage of the highest counts, or incidence in particular weeks or months, is attributed to influenza). In chapter 4 excluding even 25% of the highest P&I and ILI counts and fitting negative binomial log-link models to the remaining incidence was insufficient to account for autocorrelation and overdispersion. In this chapter, fits of 2-state Poisson HMMs to ILI and P&I incidence are described, where distinction between two states ('normal' and 'aberrant') is integrated within the modeling process. A Bayesian approach allows uncertainty about state assignments to be incorporated naturally. HMMs attribute variability in the data to more than one probability distribution and thus account for extra-Poisson variability in the data.

5.3. Data sets

Influenza years are defined as week 26 of one calendar year (the week of July 1st) to week 25 of the following calendar year. Models were fitted to P&I and ILI data by age group from the 1975/76 season to the 2004/05 season. Though ILI data are available since the 1967/68 season and P&I data since 1970/71 season, models were fitted only to data from 1975/76 since laboratory data (incorporated into multivariate models in chapter 6) are only available since 1975/76.

5.4. Description of the models

Counts are usually modeled as Poisson distributed, so Poisson log-link and identity-link HMMs were fitted (table 5.1). Log-link models allow baseline parameters - long-term trend, seasonality and artifacts - to combine multiplicatively while identity-link models allow baseline parameters to combine

additively (table 5.1). There is no *a priori* reason to exclude multiplicative or additive models since baseline parameters (long-term trend, seasonality and artifacts) may plausibly combine additively or multiplicatively. There are examples of both in the published literature (e.g. [43, 62]). Negative Binomial HMMs were also fitted where variability in the observed data in excess of that predicted by the Poisson distribution was allowed for with a separate dispersion parameter estimated from the data. As an alternative method for modeling variability in the data beyond what the Poisson HMM captures, counts were transformed using two variance-stabilising transformations - the square-root and the log - and Gaussian HMMs were then fitted to the transformed counts where overdispersion is allowed for with the separate variance parameters. Unlike the Poisson, Gaussian variance is not determined by the mean.

Table 5.1.: **Basic univariate HMMs.**

Model	**Model structure on original scale**	**Error structure on original scale**	**Handling additional overdispersion**	**Key**
Poisson log-link	multiplicative	additive	nil	LOGR
Negative Binomial log-link	multiplicative	additive	Dispersion parameter θ	NBLOGR
Poisson identity-link	additive	additive	nil	IDR
Negative Binomial identity-link	additive	additive	Dispersion parameter θ	NBIDR
Gaussian fitted to square root-transformed counts	additive (on square-root scale)	additive (on square-root scale)	separate variance parameter	SQR
Gaussian fitted to log-transformed counts	multiplicative	multiplicative	separate variance parameter	LNR

The interpretation of the mean shift, the feature of the model that distinguishes the two states, depends on the scale of analysis. In models with a log-link, the mean shift is the average ratio of a count drawn from the 'aber-

rant' state to a count drawn from the 'normal' state for a given influenza season; in models with an identity-link, the mean shift is the average difference between the count predicted from the 'aberrant' and 'normal' states for a given influenza season.

The different model structures - Poisson and Negative Binomial log-link and identity-link HMMs as well as Gaussian HMMs fitted to square-root- or log-transformed counts (table 5.1) - were fitted initially to weekly P&I counts in those aged $\geq$ 65 and ILI in those aged 15-44 to coarsely differentiate between models with different scale and distributional assumptions. These age groups were used because they contain the majority of the data in the respective data sets. Models chosen at this coarse stage were then fitted to P&I and ILI and each age group and convergence and fit were assessed in more detail. The two-state Poisson HMM with log-link had the following form:

$$
\begin{aligned}
Y_t &\sim \text{Poisson}(\mu_t) \\
\log(\mu_t)|\, S_t = 1 &= \log(N_t) + \alpha_0 + C(t, \varphi) + \beta_1 \sin\tfrac{2\pi t}{52.2} + \beta_2 \cos\tfrac{2\pi t}{52.2} \\
&+ \beta_3 \text{artifacts} \\
\log(\mu_t)|\, S_t = 2 &= \log(N_t) + \alpha_0 + \alpha_1[\text{flu season}] + C(t, \varphi) + \beta_1 \sin\tfrac{2\pi t}{52.2} + \beta_2 \cos\tfrac{2\pi t}{52.2} \\
&+ \beta_3 \text{artifacts} \\
S_t|\, S_{t-1} &\sim \text{Bernoulli}(\delta)
\end{aligned}
$$

Where Y_t is the observed P&I or ILI count in week t,

μ_t is the mean of the Poisson distribution from which Y_t is drawn,

N_t is the population offset,

α_0 is the intercept,

α_1[flu season] is the random effect yearly mean shift,

$C(t, \varphi)$ is the cubic spline with φ df,

$\beta_1 \sin \frac{2\pi t}{52.2} + \beta_2 \cos \frac{2\pi t}{52.2}$ represents seasonality, where 52.2 is the average number of weeks in a year,

$\beta_3$artifacts represents the instantaneous change in the baseline because of artifacts in the data,

and S_t is the state variable sampled from a Bernoulli distribution with probability δ, a two-by-two matrix of probabilities of moving (or not) between states at time t given which state the model was it at time t-1.

5.5. Priors

Reference priors were assigned to all parameters (table 5.2). Prior distributions are on the scale of analysis unless otherwise stated.

Table 5.2.: **Prior distributions.** Normal priors are expressed in terms of mean and precision (recall precision is the inverse of the variance).

Parameter	Prior distribution
α_0	Normal(0,0.01)
α_1[flu season]	Normal(μ_{rand}, τ_{rand})I(0,)
μ_{rand}	Normal(0,0.01)
τ_{rand}	Gamma(0.001,0.001)
β_1, β_2, β_3	Normal(0,0.01)
$\delta_{S_t=2\|S_{t-1}=1}$	Uniform(0,0.2)
$\delta_{S_t=2\|S_{t-1}=2}$	Uniform(0.6,1)
Dispersion parameter, θ[1]	Gamma(0.01,0.01)
Precision of Normal distribution[2]	Gamma(0.001,0.001)

[1]NBLOGR, NBIDR models only
[2]SQR, LNR models only

In table 5.2, $\delta_{S_t=2|S_{t-1}=1}$ is the probability of a transition into the 'aberrant' state at time t given being in the 'normal' state at time t-1, while $\delta_{S_t=2|S_{t-1}=2}$ is the probability of no transition at time t given being in the 'aberrant' state at time t-1. Prior distributions are defined as Normal(mean,precision), where the precision is the inverse of the variance, Uniform(range) and Gamma(shape,scale). I(0,) denotes restriction to positive values.

Normal and uniform reference priors were used so as to be as 'non informative' as possible, that is to approximate no prior knowledge as to the value of the parameter in question. A Bayesian analysis with reference priors used on all parameters is, in many simple situations, analogous to a frequentist analysis. The analyses described in this and the following two chapters are complex. For example, all possible sequences of states are integrated over. This means that using reference priors will not necessarily produce the same results as a frequentist analysis. Normal reference priors are equivalent to assigning approximately equal prior probability to all possible values of the parameter in question, with a weak central tendency at 0. Uniform reference priors assign equal prior probability to all possible values of the parameter within a range. Priors for transition probabilities were constrained so that the prior probability of a transition into the 'aberrant' state at time t, given being in the 'normal' state at time t-1, was uniform between 0 and 0.20, because large epidemics of influenza last approximately 8-10 weeks so a given week has an up to approximately 20% (10/52 = 0.19) chance of being the first week of a large epidemic, ignoring seasonality. The prior probability of no transition at time t, given being in the 'aberrant' state at time t-1, was uniform between 0.6 and 1 so that the state variable would 'stick' in the 'aberrant' state for the duration of the period of high incidence. [214] These priors do not allow, for example, that the probability of a transition to the 'aberrant' state at time t given having been in the 'normal' state at time t-1 is 0.21. Model results may be sensitive to these choices for priors on transition probability parameters. Given the derivation of the priors mentioned above, I think they are a reasonable place to start. Allowing for a higher than 0.2 probability of moving from the 'normal' to the 'aberrant' state in any given week and/or allowing for a lower than 0.6 probability of remaining in the 'aberrant' state from one week to the next would, if anything, increase volatility of state sequences by allow-

ing the model more freedom to move between states (see section 3.8.2). If conclusions were to be drawn from the analysis of covariate effects using the HMMs, sensitivity of results to these priors would need to be checked. Precisions and the dispersion parameter from negative binomial models were given reference Gamma distributions. [193] The prior on the mean shift was constrained to be positive to ensure that only an increase (not a decrease) in the predicted count would be labeled as arising from the 'aberrant' state.

5.6. Scale and distribution

For each parameter, two chains were run from different initial values. Models from table 5.1 fitted to P&I data for the $\geq$ 65 age group and ILI data for the 15-44 age group were run for 10,000 iterations on each of two chains (for a total of 20,000 iterations).

NBLOGR, NBIDR, SQR and LNR models were discarded in favour of simpler Poisson (LOGR and IDR) models because of poor convergence, fit or both. P&I and ILI data are underdispersed relative to the NBIDR and NBLOGR models (e.g. figure C.1). In addition, the transition probability parameters, which are key parameters for HMMs, do not converge for NBLOGR or NBIDR model fits (e.g. figure C.2).

SQR and LNR models predict P&I $\geq$ 65 data poorly (figures C.3 and C.4). In addition, coefficients from SQR models are difficult to interpret because is it not trivial to back-transform coefficients from the square-root to the original scale. [215]

LOGR and IDR models were fitted to P&I and ILI for each age group; fit and convergence of these models are summarised in the remainder of the chapter.

5.7. Model convergence

Fit and convergence of LOGR and IDR models are summarised in table 5.3.

Table 5.3.: **Summary of univariate HMM fit and convergence.** N = no; Y = yes. Volatility of the state sequence for a given model is also indicated.

Model	Outcome	Age group	LL converged?	δs converged?	State sequence converged?	Adequate fit & residuals?	Minimal residual auto-correlation?	Adequate PPD?
LOGR	P&I	0-4	N	N	N	Y	Y	Y
LOGR	P&I	5-14	Y	N	N	Y	Y	N
LOGR	P&I	15-44	Y	Y	N	Y	Y	N
LOGR	P&I	45-64	N	N	N	Y	Y	N
LOGR	P&I	$\geq$ 65	N	N	N	Y	N	Y
LOGR	ILI	0-4	Y	Y	Y, volatile	Y	Y	Y
LOGR	ILI	5-14	Y	Y	Y, volatile	Y	Y	Y
LOGR	ILI	15-44	N	Y	Y, volatile	Y	Y	Y
LOGR	ILI	45-64	Y	Y	Y	Y	Y	Y
LOGR	ILI	$\geq$ 65	Y	Y	Y	Y	Y	Y
IDR	P&I	0-4	Y	Y	N	Y	Y	Y
IDR	P&I	5-14	Y	Y	N	Y	Y	N
IDR	P&I	15-44	Y	Y	N	Y	Y	N
IDR	P&I	45-64	N	N	N	Y	N	Y
IDR	P&I	$\geq$ 65	N	Y	Y	Y	Y	Y
IDR	ILI	0-4	N	N	N	N	N	Y
IDR	ILI	5-14	N	N	N	N	N	Y
IDR	ILI	15-44	N	Y	Y	Y	Y	Y
IDR	ILI	45-64	N	N	N	N	N	Y
IDR	ILI	$\geq$ 65	N	Y	N	N	N	N

For LOGR and IDR model fits to each age group, two chains were initialised from different initial values for each parameter. Four sets of initial values were developed (one each for LOGR fits to P&I, LOGR fits to ILI, IDR fits to P&I and IDR fits to ILI). Two chains for each parameter were run for 100,000 iterations, saving 10% of sampled values for each parameter. Model fit was assessed based on plots of the final 5000 saved samples. Convergence of two chains to the same area of parameter space for key parameters was assessed using BGR plots of the LL and history plots of two chains for the LL and for the transition probability parameters. BGR plots monitor convergence of the ratio of pooled to within chains variability in the LL to 1. BGR plots also show that initial values were suitably different if the ratio of pooled to within chains variability is greater than 1 at the start of the simulation.

History plots show sampled values from every 10th iteration for two chains of the LL and the transition probabilities, starting after 20,000 iterations. Recall that apparent convergence of the two chains for the LL and transition probabilities to the same approximate value is evidence of apparent convergence of the models more generally. This is because the LL is contributed to by all parameters in the model and the transition probabilities are the defining feature of HMMs.

In general, models where transition probability parameters appear to converge show apparent convergence of most other parameters. When transition probabilities do not converge, many other parameters also do not converge (detail below).

Transition probability parameters from IDR model fits to P&I appear to converge for all but the 45-64 age group (e.g. figures 5.1 and 5.2). For IDR model fits to ILI, transition probabilities only appear to converge for fits to the 15-44 age group (e.g. figures 5.3 and 5.4). For the IDR model fit to ILI for the $\geq$ 65 age group, transition probabilities parameters appear to converge to a similar area of parameter space (figure D.1).

Transition probabilities from LOGR fits to ILI for all age groups appear to converge (e.g. figure 5.5). For LOGR fits to P&I, only the 15-44 fit shows apparent convergence (e.g. figures 5.6 and 5.7). For the LOGR model fit to P&I for the 5-14 age group, transition probability parameters appear to begin to mix in the same area of parameter space after approximately 70,000 iterations on the two chains (figure 5.8).

Apparent convergence of the LL for models mirrors apparent convergence of the transition probability parameters (e.g. figures 5.2 and 5.9) with several exceptions: the LL does not converge for IDR model fits to P&I or ILI for the ≥ 65 age group or for either LOGR or IDR fits to ILI for the 15-44 age groups, (e.g. figure D.2) though transition probabilities appeared to converge for these models (e.g. figure 5.4).

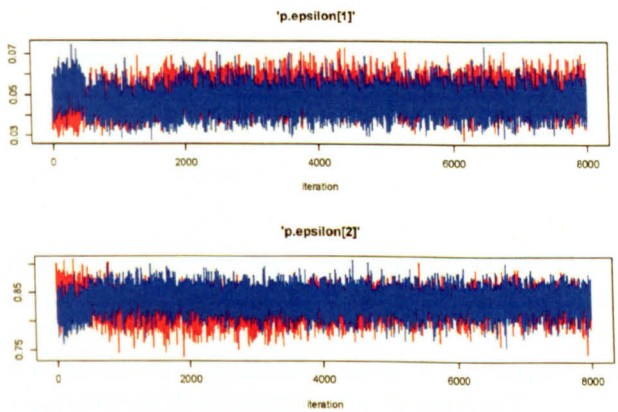


Figure 5.1.: **History plots of two chains for transition probability parameters from IDR model fit to P&I from the ≥ 65 age group.** The top plot, 'p.epsilon[1]', is two chains for $\delta_{S_t=2|S_{t-1}=1}$, the probability of a transition into the 'aberrant' state at time t given being in the 'normal' state at time t-1. The bottom plot, 'p.epsilon[2]', is two chains for $\delta_{S_t=2|S_{t-1}=2}$, the probability of no transition at time t given being in the 'aberrant' state at time t-1. Both transition probabilities appear to converge.

BGR plots of the LL for the first 20,000 iterations on two chains show that, for almost all models, chains were started from suitably disparate initial values (figures D.3 and D.4).

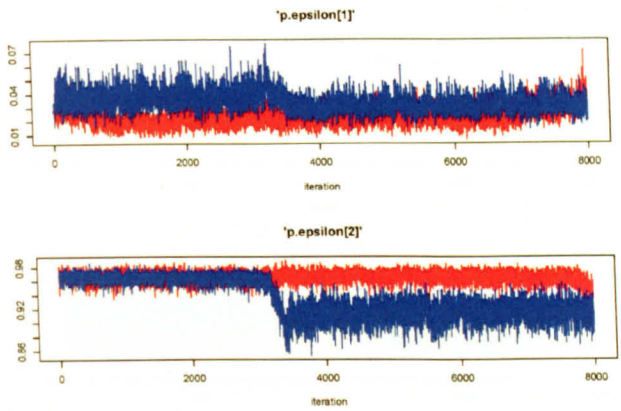


Figure 5.2.: **History plots of two chains for transition probability parameters from IDR model fit to P&I from the 45-64 age group.** Two chains for 'p.epsilon[2]' (bottom plot) diverge around the 3000th saved sample. This suggests the existence of a number of local maxima for 'p.epsilon[2]'.

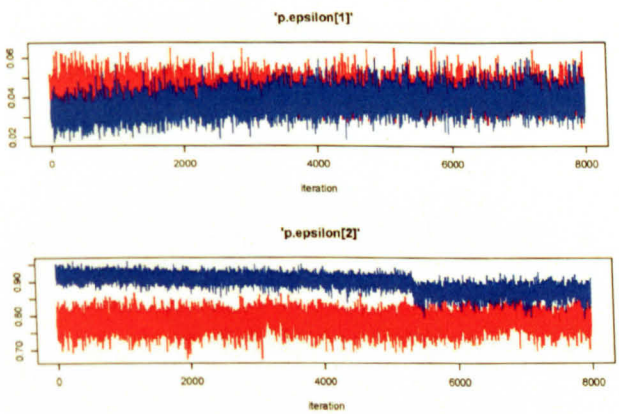


Figure 5.3.: **History plots of two chains for transition probability parameters from IDR model fit to ILI from the 5-14 age group.** 'p.epsilon[2]' (bottom plot) does not converge while 'p.epsilon[1]' appears to.

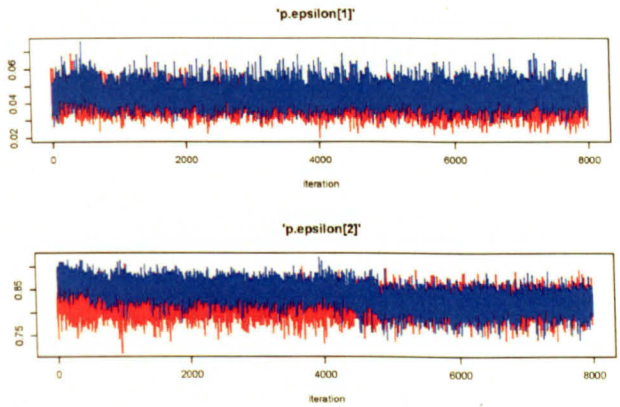


Figure 5.4.: **History plots showing apparent convergence of two chains for transition probability parameters from IDR model fit to ILI from the 15-44 age group.**

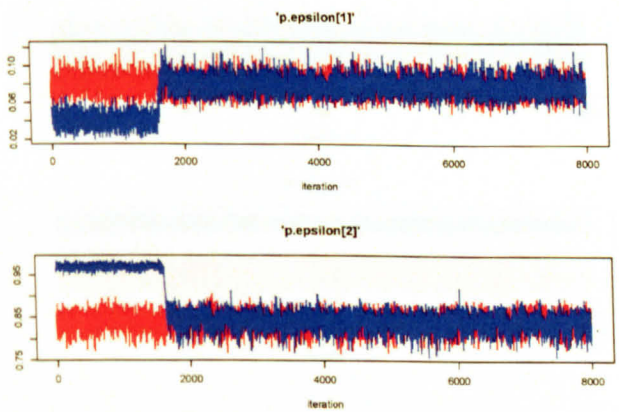


Figure 5.5.: **History plots showing apparent convergence of two chains for transition probability parameters from LOGR model fit to ILI from the 0-4 age group.**

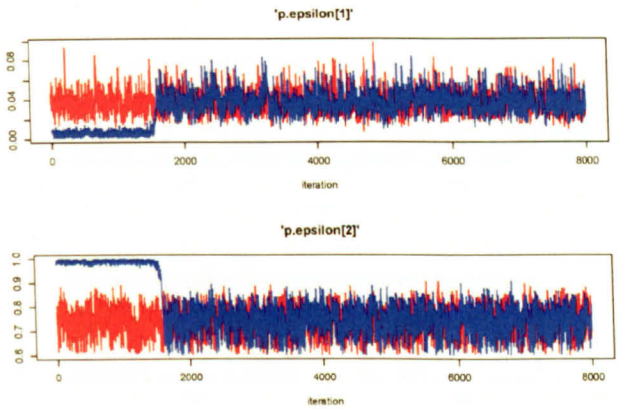


Figure 5.6.: History plots showing apparent convergence of two chains for transition probability parameters from LOGR model fit to P&I from the 15-44 age group.

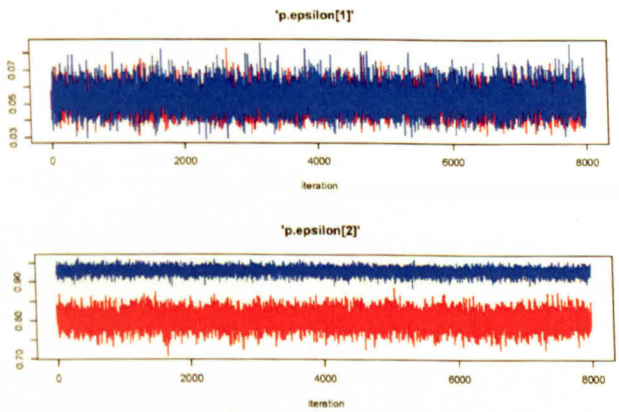


Figure 5.7.: History plots showing lack of convergence of two chains for the probability of no transition at time t given being in the 'aberrant' state at time t-1 ('p.epsilon[2]', bottom plot) from LOGR model fit to P&I from the ≥ 65 age group.

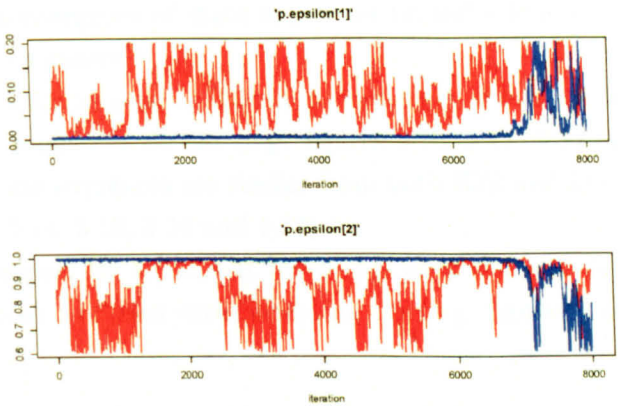


Figure 5.8.: **History plots of two chains for transition probability parameters from LOGR model fit to P&I from the 5-14 age group.** The two chains for both transition probabilities begin to inhabit the same region of the parameter space after approximately 70,000 iterations, but do not mix well.

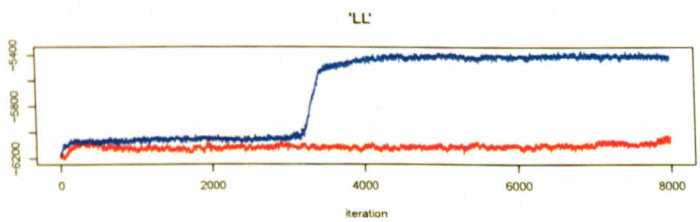


Figure 5.9.: **History plot showing lack of convergence of two chains of the LL from IDR model fit to P&I from the 45-64 age group.** Two chains for the LL diverged around the 3000th saved sample. This is consistent with the divergence for two chains for the 2nd transition probability from this model (see figure 5.2.

5.8. Convergence of state sequences

Relative goodness of fit of models was assessed by comparing state sequence plots in terms of the degree to which the two chains agreed on state assignment for each week across models, and in terms of the volatility of the state sequence. The latent, or hidden, state variables are the key parameter of HMMs and convergence of state sequences typically indicates convergence of models more generally.

For fits to P&I, in general IDR state sequences appear to converge while LOGR state sequences do not (e.g. figures 5.10, 5.11, 5.12 and 5.13). For fits to ILI, state sequences are similar from both IDR and LOGR model fits (e.g. figures 5.14, 5.15, 5.16 and 5.17).

State sequences for all LOGR fits to ILI data are volatile. Volatility of state sequences decreases with increasing age (e.g. figures 5.14 and 5.18).

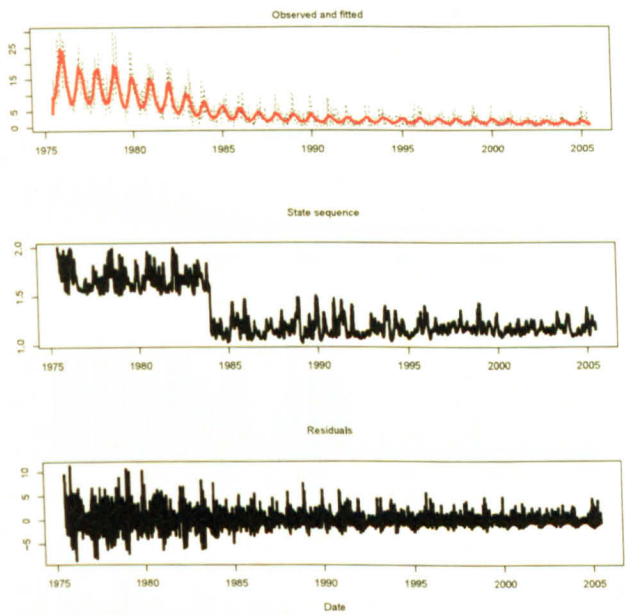


Figure 5.10.: **The state-sequence does not converge for the LOGR model fit to P&I from the 0-4 age group.** Top panel: observed (dashed) and fitted P&I data for the 0-4 age group (red); middle panel: state sequence (1.0 is the 'normal' state, 2.0 the 'aberrant' state); bottom panel: residuals (observed minus fitted P&I count for each week). The state sequence shown was plotted by averaging the state sequence estimated by each of the two chains. Lack of convergence is shown because the pooled state sequence does not clearly designate weeks as either 'normal' or 'aberrant'. This means the two chains disagreed on the state assignment for most weeks.

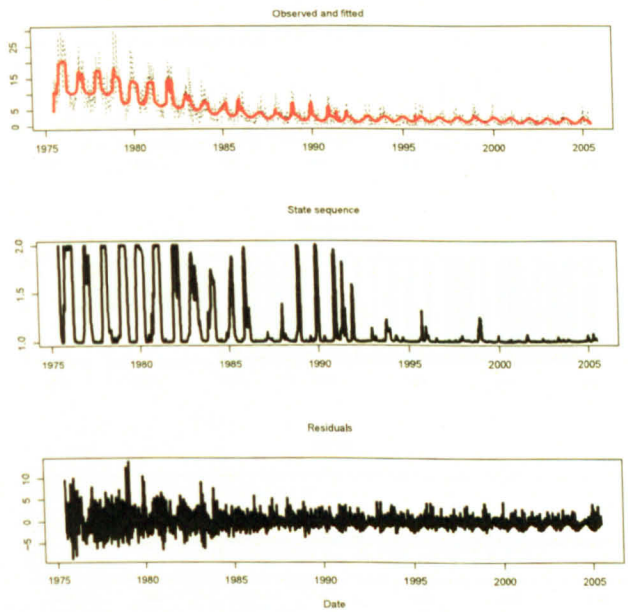


Figure 5.11.: **The IDR model (shown) is better able than the LOGR model (previous plot) to distinguish 'aberrant' from 'normal' incidence in P&I from the 0-4 age group.** The two chains appear to agree on the state assignment for most weeks.

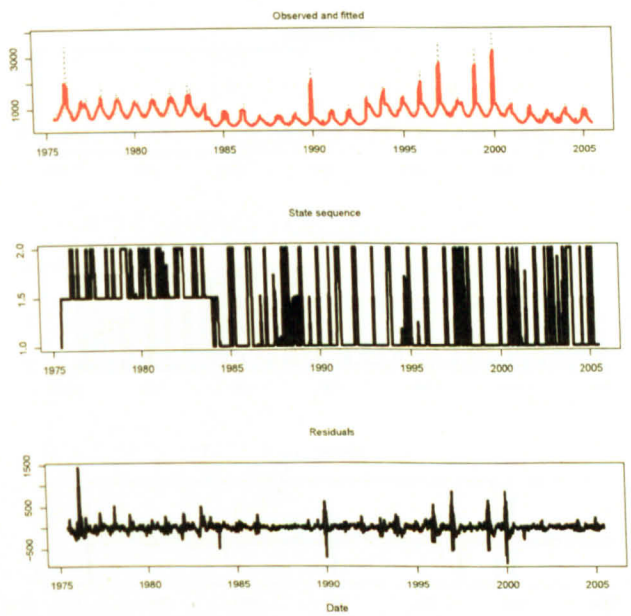


Figure 5.12.: **The state-sequence from the LOGR model fit to P&I in those $\geq$ 65 does not converge.** The two chains specifically disagree on state assignment for many weeks between 1975/76 and 1983/84.

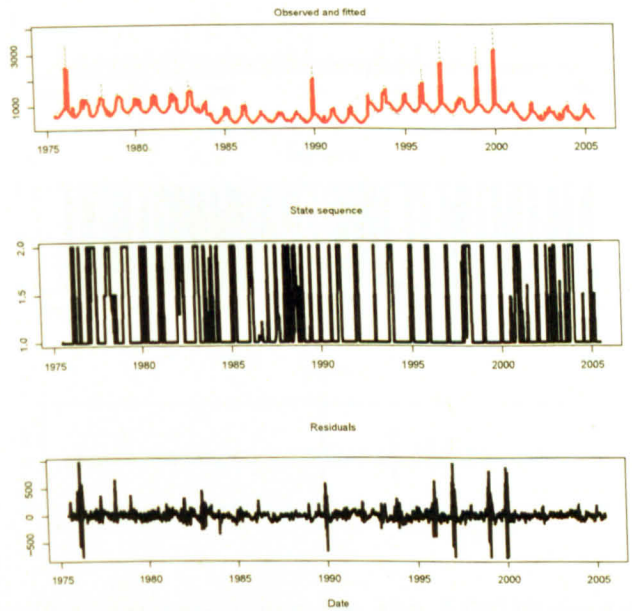


Figure 5.13.: **The state-sequence from the IDR model fit to P&I in those $\geq$ 65 appears to converge for most weeks.** Note the large residuals because of underfitting of epidemics.

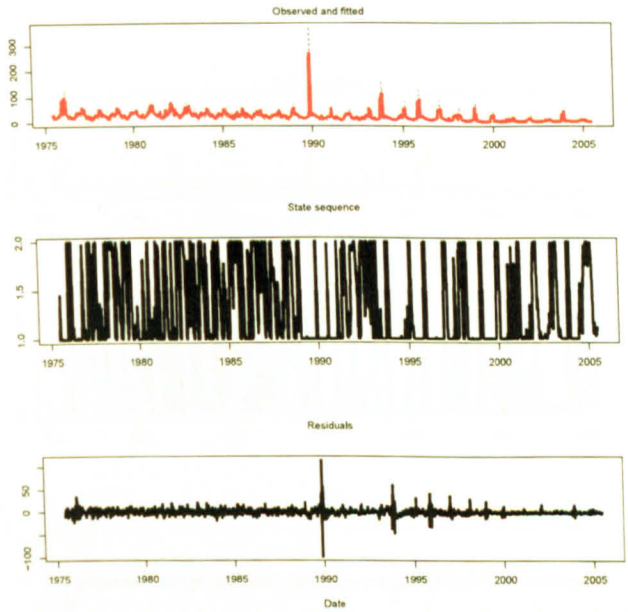


Figure 5.14.: **The state-sequence for the LOGR model fit to ILI from the 0-4 age group appears to converge for most weeks, but is volatile.** Recall that volatility of the state sequence refers to it flipping between the 'aberrant' and 'normal' states several times during influenza seasons despite the observed data indicating a single, continuous 'aberrant' period.

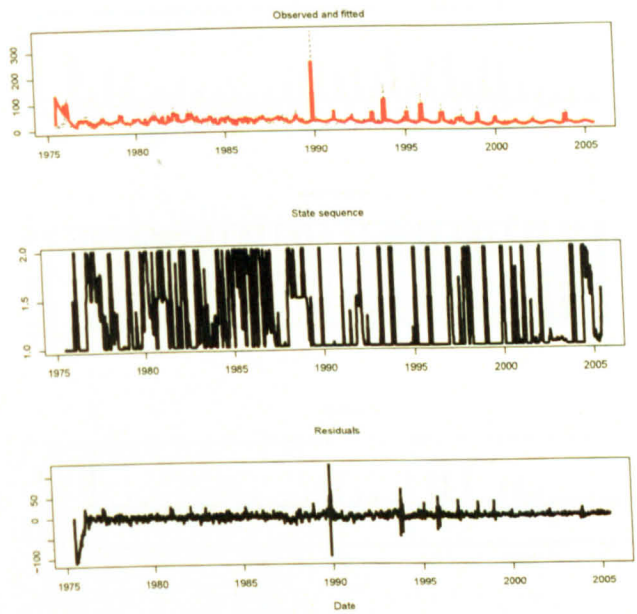


Figure 5.15.: The state-sequence for the IDR model fit to ILI in the 0-4 age group does not converge.

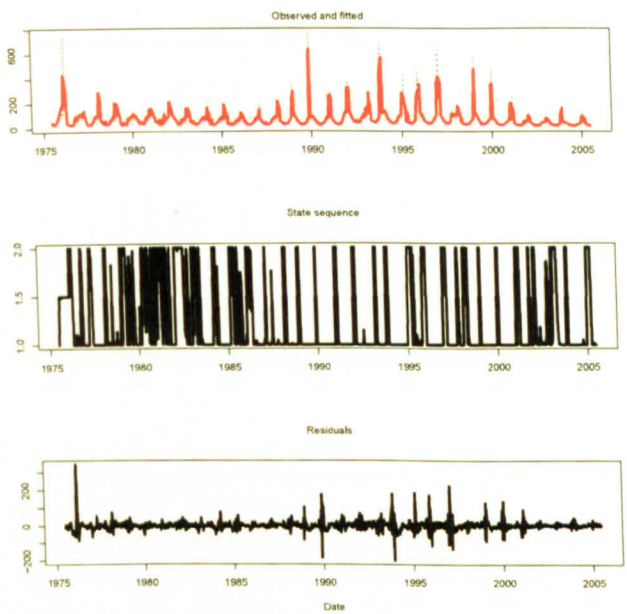


Figure 5.16.: **The state-sequence for the LOGR model fit to ILI in those 15-44 years of age appears to converge for most weeks.** There is some volatility in the state sequence.

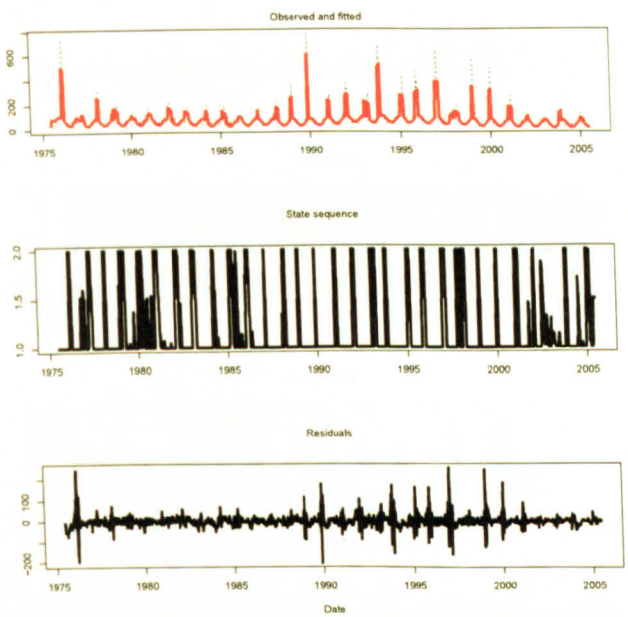


Figure 5.17.: **The state-sequence for the IDR model fit to ILI in those 15-44 years of age appears to converge.**

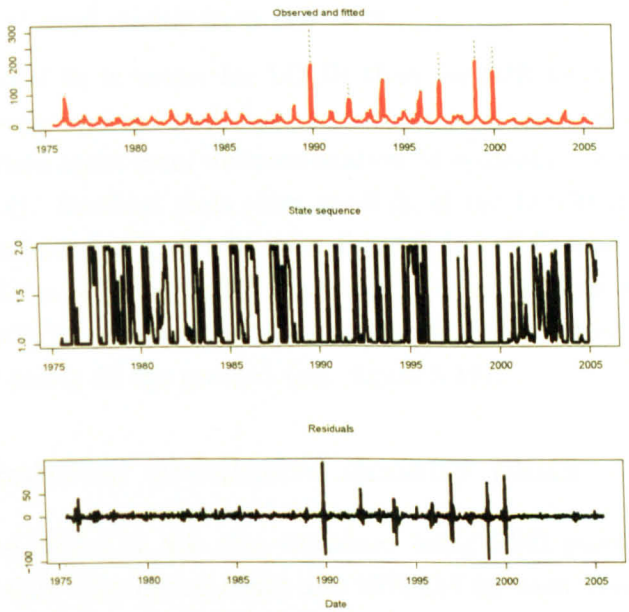


Figure 5.18.: **The state-sequence for the LOGR model fitted to ILI in those ≥ 65 (shown) appears to converge and is less volatile than the state sequences for LOGR model fits to ILI from younger age groups.** Large residuals are due to underfitting of epidemics.

5.9. Plots of observed and fitted counts and residuals

Comparing plots, averaged over two chains, of observed and fitted weekly counts and residuals between models allowed visual comparison of models with regards to model fit:

1. relative goodness of fit was evident by comparing fitted and observed counts across models to check for parts of the observed data that one or another model fits badly

2. comparing time series' of residuals across models for evidence of residual seasonality and an abundance of large residuals allowed differentiation of good models from poor ones

Overall model fit is better for LOGR than for IDR models. Observed and fitted counts and residual time series plots show good fit of all models to the P&I data apart from underestimation of epidemics (e.g. figure 5.10 in section 5.8). Residual plots show good fit of the LOGR models to ILI except for underfitting epidemics (e.g. figure 5.16). IDR models fit several periods of ILI poorly for the following age groups: the first season in the data set (1975/76 - all but the 15-44 age group); the period from 1986 to 1988 (45-64 and $\geq$ 65 age groups) (e.g. figure 5.19).

5.10. Posterior predictive density plots

Comparative model fit was also visualised using PPD plots. Recall that PPD plots show a predicted count and 95% CrI for each week drawn from the sampling distribution for the observed data. P&I data from the 0-4, 5-14 and 15-44 age groups are underdispersed relative to both the LOGR and IDR models (data not shown). P&I from the 45-64 age group are underdispersed relative to the LOGR model but adequately modeled by the IDR model (figure D.5). P&I for the $\geq$ 65 are neither overdispersed nor underdispersed relative to LOGR and IDR models (figure 5.20).

ILI for all age groups are neither overdispersed nor underdispersed relative to LOGR (e.g. figure D.6). For IDR model fits to ILI for the 0-4, 5-14, 45-64 and $\geq$ 65 age group, posterior predictive CrIs are very large for the 1975/76 and 2004/05 influenza seasons (also shown in figure D.6).

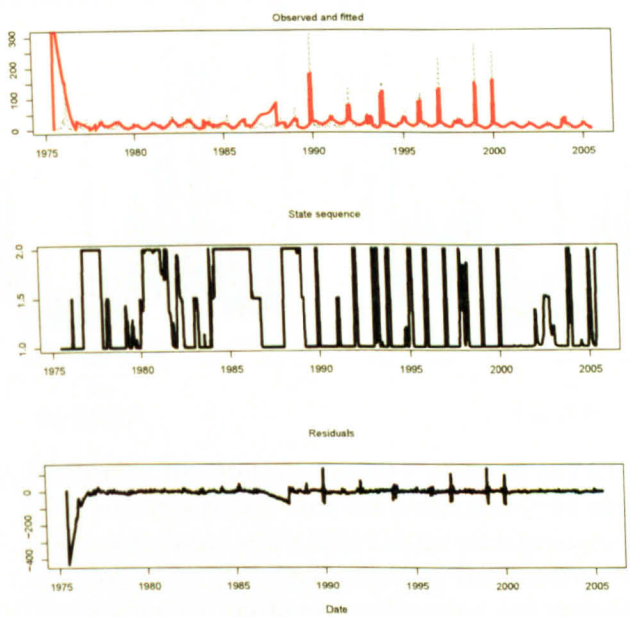
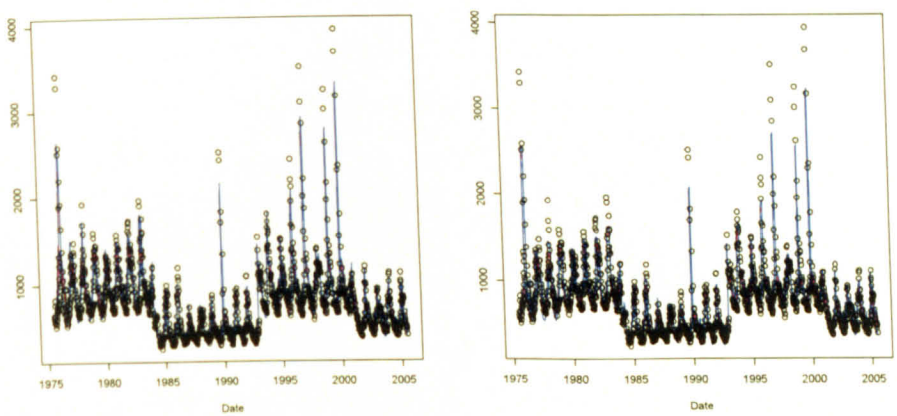


Figure 5.19.: **IDR model-predicted ILI for the $\geq$ 65 age group (top plot) are very different from observed ILI for the 1975/76 influenza seasons and for the period between 1986 and 1988.**



(a) LOGR (b) IDR

Figure 5.20.: **Posterior predictive density plots of (a) LOGR and (b) IDR models fitted to P&I from the $\geq$ 65 age group.** The predicted counts and a 95% CrI for each predicted count (lines) are plotted on the same graph as the observed data (circles). These data are neither overdispersed nor underdispersed relative to the LOGR and IDR models and could be adequately modeled by either. Recall that overdispersion is present if many observed data fall outside the posterior predictive CrIs. Underdispersion is present if very few or no observed data fall outside predicted CrIs.

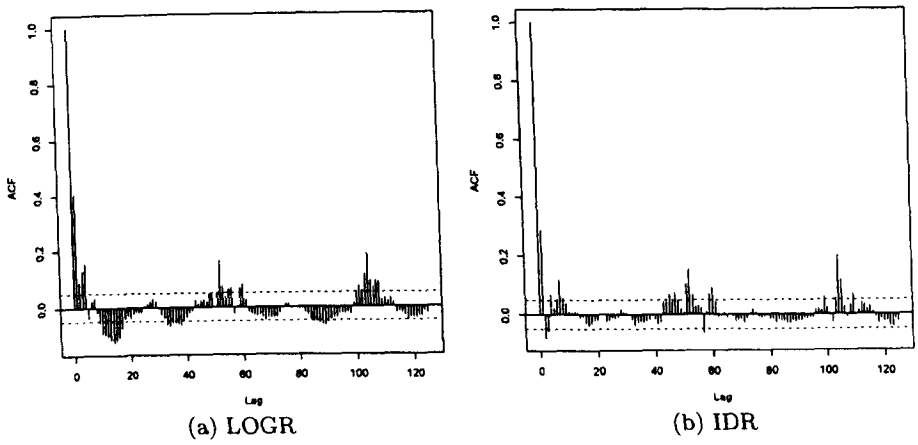
5.11. Autocorrelation in residuals

Observations in a time series of counts of a seasonal infectious disease are not independent; counts at short lags (e.g. 1-2 weeks) are correlated because of person-to-person transmission and counts from the similar weeks in different years are correlated because of seasonality (since the influenza season occurs around the same time each year). In the models described in this and the following two chapters, the Markov chain generates autocorrelation. Models described in each results chapter have used a Fourier term (one sine and one cosine term) and cubic splines to account for seasonality (and long-term trend) in the data. The degree to which different models account for the correlation between counts at lags of 1-2 weeks and at lags of around 52 weeks was assessed by plotting correlation of residuals at lags up to 120 weeks. Recall that horizontal dotted lines indicate the threshold below which autocorrelation is 'ignorable' ($\pm 2/\sqrt{1566}$). [204]

Autocorrelation plots from both LOGR and IDR model fits to P&I for the $\geq$ 65 age group show underfitting of epidemics (evidenced by positive autocorrelation at short lags) (figure 5.21). They also show inadequate modeling of seasonality since there is positive autocorrelation at lags of approximately 52 and 104 weeks.

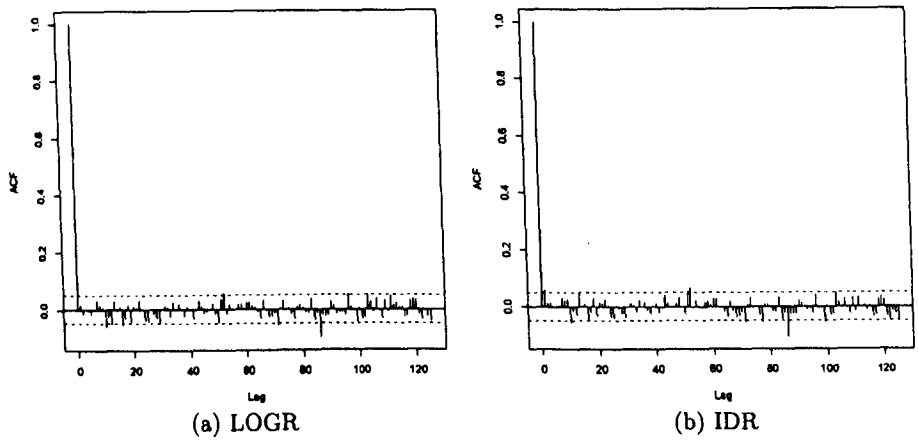
The IDR model badly fits the P&I data for the 45-64 age group (figure D.7). These data are adequately modeled by the LOGR model. There is minimal autocorrelation in residuals from LOGR or IDR model fits to P&I data for the 0-4, 5-14 and 15-44 age groups (e.g. figure 5.22).

Residuals from IDR model fits to ILI are generally highly autocorrelated for many consecutive weeks, suggesting poor overall model fit (e.g. figure 5.23b). Residuals from LOGR model fits to ILI are generally less highly correlated than IDR fits. For the LOGR model fit to ILI in the $\geq$ 65 age group there is a large negative correlation at a lag of 2 weeks (figure 5.23a).



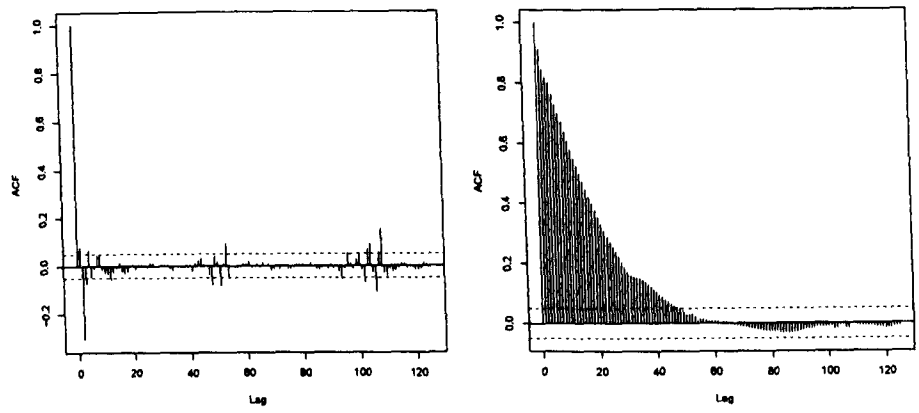
(a) LOGR (b) IDR

Figure 5.21.: **Autocorrelation plots of residuals from (a) LOGR and (b) IDR model fits to P&I from the $\geq$ 65 age group.** The correlation between residuals against the lag between the residuals. Correlation between residuals is 1 at lag 0 because this correlation is between the residual and itself. Horizontal dotted lines are set at $\pm 2/\sqrt{1566}$. Both models underfit epidemics (shown by positive autocorrelation at lag 1 week). Both models also inadequately model seasonality (note positive autocorrelation at lags of approximately 52 and 104 weeks).



(a) LOGR (b) IDR

Figure 5.22.: **There is minimal autocorrelation in residuals from (a) LOGR and (b) IDR model fits to P&I from the 15-44 age group.**



(a) LOGR (b) IDR

Figure 5.23.: **Autocorrelation plots of residuals from (a) LOGR and (b) IDR model fits to ILI from the $\geq$ 65 age group.** The LOGR model fit shows minimal autocorrelation apart from a large negative correlation at a lag of 2 weeks. The IDR model fit is highly autocorrelated for many consecutive weeks showing poor overall model fit.

5.12. Relationship between information and convergence

Lack of convergence of transition probabilities (and thus state sequences) for some models may reflect a lack of information in P&I and ILI about state transitions and thus mean shifts. For example, two chains for transition probabilities from the LOGR fit to P&I for the 0-4 age group do not easily explore the parameter space. Instead, the two chains wander almost at random across the parameter space and the state sequence does not converge (figures 5.10 and D.8). While in principle if there is little information in the data to inform the posterior for transition probabilities the posterior should look like the prior (i.e. diffuse), in practice exploring a complex parameter space is difficult when the data do not contain information to guide the sampler to the appropriate part of the parameter space. This issue is explored further in chapter 6 in the context of multivariate models and the sharing of information about influenza across P&I, ILI and laboratory data for influenza A. It is revisited in chapter 7 in the context of informative priors on the effect of CT seasons and impact of vaccination on the mean shift.

5.13. Summary of results

LOGR and IDR models are chosen and negative binomial and Gaussian models discarded, because of problems of convergence and model fit with negative binomial and Gaussian models. It may be that the negative binomial and Gaussian HMMs fail to distinguish overdispersion relative to the Poisson distribution from variability generated by the hidden Markov process (see section 8.5.2).

When modeling P&I, generally IDR models appear to converge and LOGR models do not. When modeling ILI, LOGR models are preferred. When the state sequence does not converge it is because the transition probability parameters for that model do not converge and vis versa. P&I for all age groups apart from the ≥ 65 age group are underdispersed relative to both LOGR and IDR models. ILI are not underdispersed or overdispersed relative to either LOGR or IDR models. Both LOGR and IDR models are taken forward for joint modeling (chapter 6). Values of particular coefficients (e.g.

for the mean shifts) are not interpreted until chapter 7.

5.14. Strengths of univariate HMMs

The Poisson HMMs presented adequately model variability in the P&I and ILI data; overdispersion is no longer an issue. By allowing the model to determine whether the data are consistent with there being two probability distributions underlying observed counts there is no need to designate 'aberrant' from 'normal' weeks externally to model fitting. Uncertainty in the distinction between 'normal' and 'aberrant' state is incorporated naturally.

5.15. Limitations

As discussed in chapter 3, there are several limitations to the data modeled which may explain some modeling problems and which motivate joint models in chapter 6.

P&I and ILI have variable specificity for influenza. Estimation of the impact of vaccination and of CT seasons - where any impact is expected to be restricted to outcomes caused by or attributable to influenza - necessitates increasing specificity of the P&I and ILI data for influenza. Low specificity might explain difficulties in clearly estimating state sequences. This limitation is addressed in chapter 6 where multivariate models are fitted to P&I and laboratory reports for influenza A, to ILI and laboratory reports, or to all three. Since laboratory-confirmed incidence has high specificity for influenza, jointly estimating state sequences using P&I/ILI data and laboratory data increases specificity of state sequences for influenza.

A limitation of univariate HMMs is that model fit to epidemics is poor for all models. Some of this underestimation is because key explanatory factors, influenza A/H3N2 dominance, vaccination and CT seasons as examples, are not included in these models. Also, recall that for LOGR models the mean shift for each influenza season is the average ratio of incidence during 'aberrant' periods to 'normal' periods for that season. The underfitting of epidemics might explain autocorrelation of residuals at short lags. In chapter 7, the value of these factors for explaining variability in mean shifts between seasons is explored.

The approach to modeling seasonality, with one sine and one cosine term,

as introduced by Serfling and adapted by many authors [70, 81, 92], is limited by its rigidity. In the models fitted in the thesis, cubic splines with 5-14 df used to capture long-term trend and dummy variables controlling for artifacts in 'normal' incidence increase the flexibility of the HMM to capture seasonality (and long-term trend) in the data relative to the sine and cosine term alone. However, autocorrelation plots of residuals show there is inadequate modeling of seasonality. Future work could include modeling seasonality more flexibly. Alternatives to a Fourier term which offer increased flexibility for modeling seasonality are using indicator variables for month (e.g. [52]) or splines with several df each year (e.g. [42]).

6. Joint HMM results

6.1. Aims of this chapter

The aim of the work described in this chapter was to increase specificity of P&I and ILI for influenza by fitting multivariate HMMs, where the state sequence is estimated jointly across age-specific counts of laboratory-confirmed influenza A cases along with ILI and/or P&I, to determine models to which CT seasons, the antigenic distance between clusters and vaccine coverage could be added in chapter 7.

6.1.1. Objectives of this chapter

1. To explore model fit and convergence of age group-specific bivariate models (fitted first to laboratory and ILI data together and then to laboratory and P&I data together) and trivariate models (modeling laboratory, P&I and ILI together in a single model) in both the Poisson log-link and Poisson identity-link framework

2. To look for evidence for lagged state transitions between outcome variables within an age group by overlaying state sequences from univariate LOGR and IDR fits to P&I, ILI and laboratory reports for influenza A

3. To explore fitting multivariate models that allow for a lag between the effect of the state sequence on P&I/ILI relative to its effect on laboratory reports for influenza A

6.1.2. Main findings

In general, multivariate LOGR models are preferred to multivariate IDR models because most of the multivariate LOGR models appear to converge and to fit the data better than multivariate IDR models. Modeling P&I

or ILI jointly with laboratory reports increases the precision of the random effect mean shift compared to univariate models in most cases. There is evidence of a lag between the timing of 'aberrant' periods in different outcome variables for a given age group. The lag in the timing of 'aberrant' periods varies across influenza seasons. Models allowing for a constant lag in the timing of 'aberrant' periods across outcome variables are computationally difficult and are not developed beyond showing that OpenBUGS can fit these models in principle. Multivariate LOGR models were taken forward for estimating the effect of cluster transitions and the impact of vaccination on the mean shift in chapter 7. Multivariate IDR models were discarded because of lack of convergence and poor model fit.

6.2. Introduction

In chapter 5, LOGR models were shown to fit ILI but not P&I and IDR models were shown to fit P&I but to be worse at fitting ILI than LOGR models were. In this chapter, multivariate LOGR and IDR models were fitted and improvement in convergence and model fit compared to univariate models was explored. Multivariate models appropriate for estimating the effect of cluster transitions and the impact of vaccination were decided upon.

6.3. Data sets

In this chapter, P&I and ILI data were modeled jointly with laboratory reports for influenza A. Laboratory reports for influenza A were used instead of laboratory reports for influenza A and B, combined, because there is evidence that influenza years dominated by circulation of influenza A/H3N2 virus experience higher peak incidence of lab/clinical incidence. [92] Influenza A therefore contributes more than influenza B to the variability in the impact of influenza seasons. Laboratory data provided by the HPA CfI do not include subtype (A/H3N2, A/H1N1).

6.4. Description of the model

Age group-specific bivariate and trivariate LOGR and IDR models were fitted (table 6.1). Bivariate HMMs were fitted first to weekly counts of P&I

and laboratory reports for influenza A and second to ILI and laboratory reports. Trivariate models were fitted to P&I, ILI and laboratory reports, simultaneously. Figure 6.1 shows a schematic of the bivariate HMM where t denotes 1 week and arrows denote conditional dependencies.

Table 6.1.: **Joint HMMs.**

Model	Key
Bivariate Poisson log-link	biLOGR
Trivariate Poisson log-link	triLOGR
Bivariate Poisson identity-link	biIDR
Trivariate Poisson identity-link	triIDR

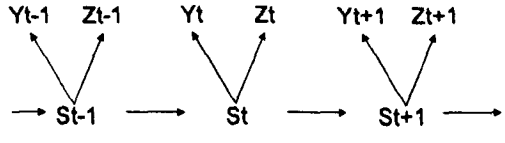


Figure 6.1.: **Schematic of a bivariate HMM.**

In multivariate HMMs, state sequences were jointly estimated by the two or three outcome variables in each model. An assumption inherent in these models is that state transitions occur in the same weeks across outcome variables. This assumption is explored in more detail in section 6.12. Below is the formula for the biLOGR HMM.

$$Y_{P\&I_t} \sim \text{Poisson}(\mu_{P\&I_t})$$

$$Y_{lab_t} \sim \text{Poisson}(\mu_{lab_t})$$

$$\begin{aligned}\log(\mu_{P\&I_t})|\ S_t = 1 &= \log(N_t) + \alpha_{0_{P\&I}} + C_{P\&I}(t, \varphi_{P\&I}) \\ &+ \beta_{1_{P\&I}} \sin\tfrac{2\pi t}{52.2} + \beta_{2_{P\&I}} \cos\tfrac{2\pi t}{52.2} \\ &+ \beta_{3_{P\&I}} \text{artifacts}\end{aligned}$$

$$\begin{aligned}\log(\mu_{P\&I_t})|\ S_t = 2 &= \log(N_t) + \alpha_{0_{P\&I}} + \alpha_1[\text{flu season}] + C_{P\&I}(t, \varphi_{P\&I}) \\ &+ \beta_{1_{P\&I}} \sin\tfrac{2\pi t}{52.2} + \beta_{2_{P\&I}} \cos\tfrac{2\pi t}{52.2} \\ &+ \beta_{3_{P\&I}} \text{artifacts}\end{aligned}$$

$$\begin{aligned}\log(\mu_{lab_t})|\ S_t = 1 &= \log(N_t) + \alpha_{0_{lab}} + C_{lab}(t, \varphi_{lab}) \\ &+ \beta_{1_{lab}} \sin\tfrac{2\pi t}{52.2} + \beta_{2_{lab}} \cos\tfrac{2\pi t}{52.2} \\ &+ \beta_{3_{lab}} \text{artifacts}\end{aligned}$$

$$\begin{aligned}\log(\mu_{lab_t})|\ S_t = 2 &= \log(N_t) + \alpha_{0_{lab}} + \alpha_1[\text{flu season}] + C_{lab}(t, \varphi_{lab}) \\ &+ \beta_{1_{lab}} \sin\tfrac{2\pi t}{52.2} + \beta_{2_{lab}} \cos\tfrac{2\pi t}{52.2} \\ &+ \beta_{3_{lab}} \text{artifacts}\end{aligned}$$

$$S_t|\ S_{t-1} \sim \text{Bernoulli}(\delta)$$

Where Y_t is the observed count of P&I or laboratory reports for influenza A in week t,
μ_t are the respective means of the Poisson distributions from which Y_t are drawn,
N_t is the population offset,
α_0 are the intercepts,
α_1[flu season] are the random effect yearly mean shifts,
$C(t, \varphi)$ are the cubic splines with φ df,
$\beta_1 \sin \frac{2\pi t}{52.2} + \beta_2 \cos \frac{2\pi t}{52.2}$ represent seasonality, where 52.2 is the average number of weeks in a year,
$\beta_3$artifacts represent the instantaneous change in the baseline because of artifacts in the data,
and S_t is the state variable sampled from a Bernoulli distribution with probability δ, where δ is a two-by-two matrix of probabilities of moving between states at time t given which state the model was it at time t-1, jointly estimated between the P&I and laboratory data in this case.

Multivariate IDR models had the same form as the multivariate LOGR models except that in multivariate IDR models μ, not $\log(\mu)$, was dependent on the linear predictor and there was no population offset.

6.5. Priors

All parameters were given reference priors as in the previous chapter (table 5.2)

6.6. Model convergence

Fit and convergence of bivariate models are summarised in table 6.2 while fit and convergence of trivariate models are summarised in table 6.3.

Table 6.2.: **Summary of model fit and convergence: bivariate HMMs.** N = no; Y = yes. Volatility of the state sequence for a given model is also indicated. When none of the LL, transition probability parameters (δs) or state sequences appeared to converge, model fit was not assessed (blank fields in the table).

Model	Outcome	Age group	LL converged?	δs converged?	State sequence converged?	Adequate fit & residuals?	Minimal residual auto-correlation?	Adequate PPD?
biLOGR	P&I	0-4	N	Y	Y	Y	Y	Y
biLOGR	P&I	5-14	Y	Y	Y	Y	Y	N
biLOGR	P&I	15-44	Y	Y	Y	Y	Y	Y
biLOGR	P&I	45-64	Y	Y	Y	Y	Y	Y
biLOGR	P&I	$\geq$ 65	N	N	N			
biLOGR	ILI	0-4	Y	Y	Y	Y	Y	Y
biLOGR	ILI	5-14	Y	Y	Y, volatile	Y	Y	Y
biLOGR	ILI	15-44	Y	Y	Y, volatile	Y	Y	Y
biLOGR	ILI	45-64	Y	Y	Y	Y	Y	Y
biLOGR	ILI	$\geq$ 65	Y	Y	Y	Y	Y	Y
biIDR	P&I	0-4	Y	Y	Y	N	N	Y
biIDR	P&I	5-14	Y	Y	Y	N	Y	Y
biIDR	P&I	15-44	Y	Y	Y	Y	Y	Y
biIDR	P&I	45-64	Y	Y	Y	Y	Y	Y
biIDR	P&I	$\geq$ 65	N	N	N			
biIDR	ILI	0-4	N	N	N			
biIDR	ILI	5-14	N	N	N			
biIDR	ILI	15-44	Y	Y	Y	Y	Y	Y
biIDR	ILI	45-64	N	N	N			
biIDR	ILI	$\geq$ 65	N	N	N			

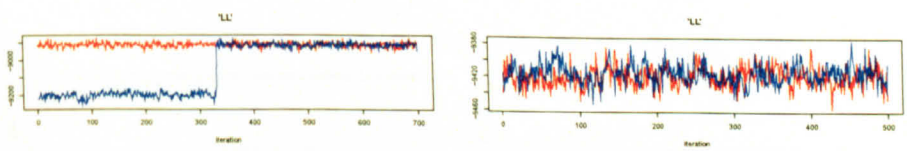
Table 6.3.: **Summary of model fit and convergence: trivariate HMMs.** Blanks in the LL, δs and state sequence columns reflect that triLOGR models were fitted jointly to P&I, ILI (and laboratory) data so LL, δs and state sequences are shared between P&I and ILI for a given model and age group.

Model	Outcome	Age group	LL converged?	δs converged?	State sequence converged?	Adequate fit & residuals?	Minimal residual auto-correlation?	Adequate PPD?
triLOGR	P&I	0-4	Y	Y	Y	Y	Y	N
triLOGR	ILI	0-4				Y	Y	Y
triLOGR	P&I	5-14	Y	Y	Y, volatile	Y	Y	N
triLOGR	ILI	5-14				Y	Y	Y
triLOGR	P&I	15-44	Y	Y	Y	Y	Y	Y
triLOGR	ILI	15-44				Y	Y	Y
triLOGR	P&I	45-64	Y	Y	Y	Y	Y	Y
triLOGR	ILI	45-64				Y	Y	Y
triLOGR	P&I	≥ 65	N	N	N			
triLOGR	ILI	≥ 65						
triIDR	P&I	0-4	N	Y	Y	N	N	Y
triIDR	ILI	0-4				N	N	Y
triIDR	P&I	5-14	N	Y	Y	N	Y	Y
triIDR	ILI	5-14				N	N	Y
triIDR	P&I	15-44	Y	Y	Y	Y	Y	Y
triIDR	ILI	15-44				Y	Y	Y
triIDR	P&I	45-64	N	N	Y	Y	Y	Y
triIDR	ILI	45-64				N	Y	Y
triIDR	P&I	≥ 65	N	N	Y	Y	Y	Y
triIDR	ILI	≥ 65				N	N	Y

Two chains for each parameter were started from different sets of initial values and were run for 100,000 iterations, saving 1% of sampled values. Saving only 1% of sampled values was done to ease storage and handling of the large files generated by the analyses. Despite this high degree of thinning, Monte Carlo error (MC-error), an estimate of the difference between the mean of the sampled values and the true posterior mean, is less than 5% of the standard deviation of posterior distributions for most parameters from all models. This is an indication that posterior estimates are sufficiently accurate. Six sets of initial values were developed (one each for biLOGR fits to P&I, biLOGR fits to ILI, biIDR fits to P&I, biIDR fits to ILI, triLOGR and triIDR models). Model fit was assessed based on plots of the final 500 saved samples. History plots of the LL and transition probabilities were used to assess model convergence. BGR plots of the LL for the first 20,000 iterations were used to show initial values were disparate enough.

For bivariate fits to P&I and laboratory reports, most model LLs appear to converge (e.g. figure 6.2). For bivariate fits to ILI and laboratory reports, only biLOGR LLs appear to converge (e.g. figure 6.3).

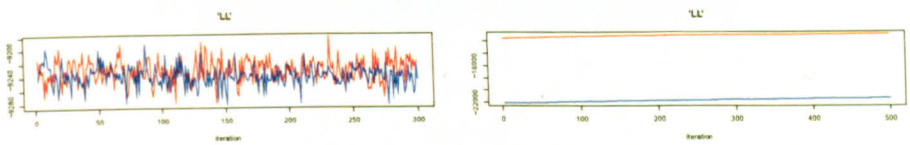
From trivariate model fits, the LL from triLOGR models fitted to the 0-4, 5-14, 15-44 and 45-64 age groups appear to converge (e.g. figure 6.4); the LL from most triIDR models does not converge (also shown in figure 6.4).



(a) biLOGR (b) biIDR

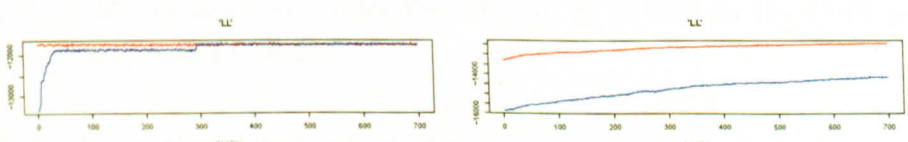
Figure 6.2.: **Example history plots showing apparent convergence of the LL from (a) biLOGR and (b) biIDR models fitted to P&I.** This plot shows fits to P&I from the 45-64 age group.

Apparent convergence of the transition probability parameters roughly parallels apparent convergence of the LL for most bivariate models. For the biLOGR fit to P&I for the 0-4 age group, transition probabilities appear to converge despite the LL not converging (figures 6.5 and E.1). For the triIDR model fits to the 45-64 and $\geq$ 65 age groups, two chains for the



(a) biLOGR (b) biIDR

Figure 6.3.: **Example history plots showing (a) apparent convergence of the LL from biLOGR fits to ILI and (b) lack of convergence of the LL from biIDR models fitted to ILI.** This plot shows fits to ILI from the ≥ 65 age group.



(a) triLOGR (b) triIDR

Figure 6.4.: **Example history plots showing (a) apparent convergence of the LL from triLOGR models and (b) lack of convergence of the LL from triIDR models.** This plot shows fits to data for the 0-4 age group.

transition probabilities appear to converge to similar areas of the parameter space (e.g. figure 6.6).

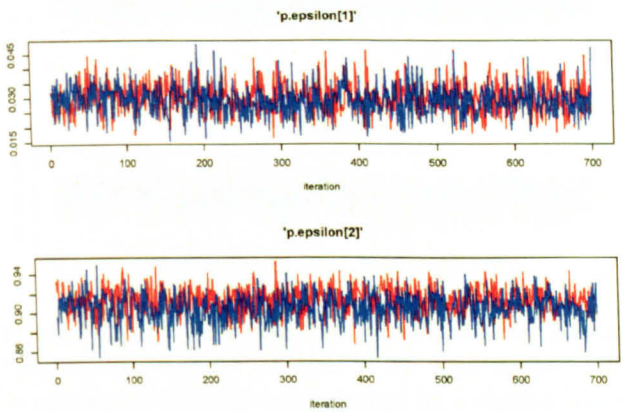


Figure 6.5.: **History plot showing apparent convergence of transition probability parameters from the biLOGR model fit to P&I and laboratory reports from the 0-4 age group.** Recall that 'p.epsilon[1]', the top panel, refers to $\delta_{S_t=2|S_{t-1}=1}$, the probability of a transition into the 'aberrant' state at time t given being in the 'normal' state at time t-1. The bottom panel, 'p.epsilon[2]', refers to $\delta_{S_t=2|S_{t-1}=2}$, the probability of no transition at time t given being in the 'aberrant' state at time t-1.

BGR plots of the LL for the first 20,000 iterations show that initial values are suitably disparate except for the biLOGR fit to P&I for the 15-44 age group (figures E.2 to E.5).

6.7. Convergence of state sequences

6.7.1. Models fitted to P&I and laboratory data

For fits to P&I, state sequences appear to converge for biLOGR model fits to the 0-4 and 5-14 age groups (e.g. figure 6.7), when they do not converge for (univariate) LOGR model fits (e.g. figure 5.10). This suggests one of two possible explanations. There may be sharing of information across outcomes in the biLOGR models to increase power to estimate the state sequence. Alternatively, laboratory data may dominate the estimation of the transition probabilities and state sequences in the biLOGR model

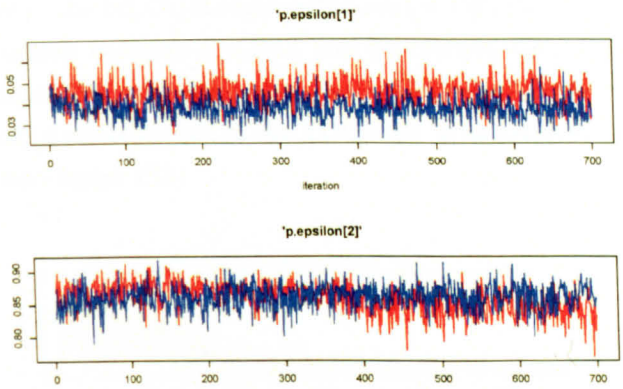


Figure 6.6.: **History plot of transition probability parameters from triIDR model fit to the $\geq$ 65 age group.** Two chains for 'p.epsilon[1]' appear to converge to a similar area of parameter space though the two chains do not lie on top of one another.

fits. Recall the relationship between lack of information and convergence introduced in section 5.12. The state sequences for univariate LOGR and IDR model fits to laboratory reports for influenza A for the 0-4 and 5-14 age groups appear to converge by 20,000 model iterations, meaning that the univariate models readily recognise the laboratory data to be consistent with two states (figures F.1, F.2, F.3 and F.4). This means that there is ample information in the laboratory data to allow the model to differentiate two states ('aberrant' and 'normal') within them.

State sequences from biIDR models fitted to P&I for the 0-4 and 5-14 age groups appear to converge as in the univariate case (e.g. figures 6.8 and 5.11).

State sequences from biLOGR and biIDR fits to P&I from the 15-44 and 45-64 age groups appear to converge, as they do in the univariate case (e.g. figures 6.9 and E.7).

Neither biLOGR nor biIDR fits to P&I for the $\geq$ 65 age group appear to converge (figures 6.10 and 6.11). The state sequence from the biLOGR fit to P&I for the $\geq$ 65 age group is similar to that from the (univariate) LOGR fit to P&I for the $\geq$ 65 age group (figure 5.12). The biIDR state sequence from the fit to P&I for the $\geq$ 65 age group is worse than the IDR fit to P&I for the $\geq$ 65 age group (figure 5.13). Worse convergence of the biIDR compared to IDR model fits to P&I in the $\geq$ 65 age group, and lack

of convergence of the biLOGR state sequence for P&I in the $\geq$ 65 age group, may reflect conflict between the P&I and laboratory data for the $\geq$ 65 age group in estimating a joint state sequence. This pattern of results may also reflect the lack of convergence of the LOGR fit to laboratory reports for the $\geq$ 65 age group (figure F.5).

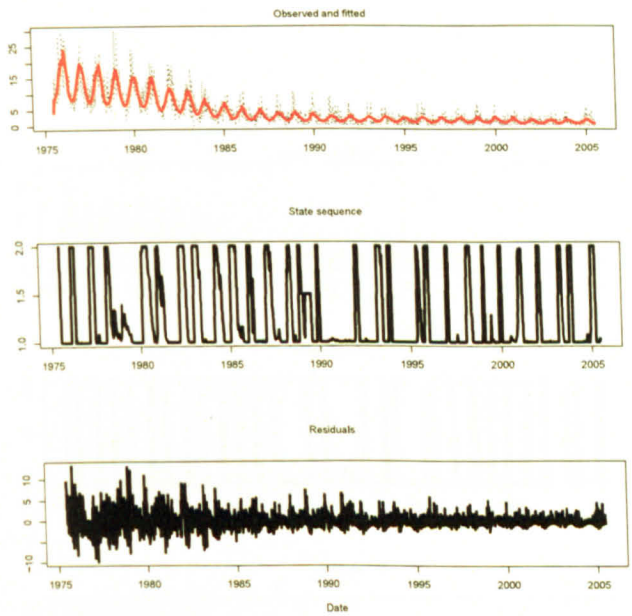


Figure 6.7.: **The state-sequence for biLOGR model fit to P&I and laboratory reports in those 0-4 years of age.** The state sequence appears to converge for most seasons and is clearly estimated apart from the 1978/79 and 1988/89 seasons where the two chains disagree on state assignment for some weeks. Recall that the state sequence shown is the average of the state sequences estimated by each of the two chains. Top panel: observed (dashed) and fitted P&I data (red); middle panel: state sequence jointly estimated between the P&I and laboratory data (1.0 is the 'normal' state, 2.0 the 'aberrant' state); bottom panel: residuals (observed minus fitted P&I count for each week).

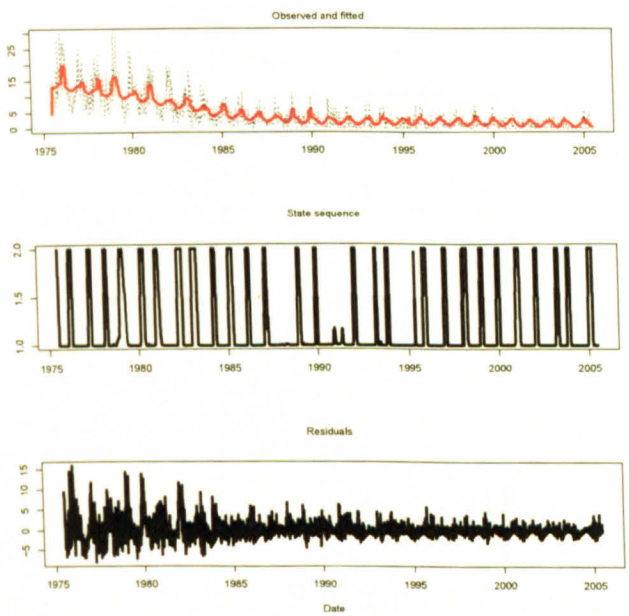


Figure 6.8.: The state-sequence for the biIDR model fitted to P&I and laboratory reports in the 0-4 age group appears to converge and is clearly estimated.

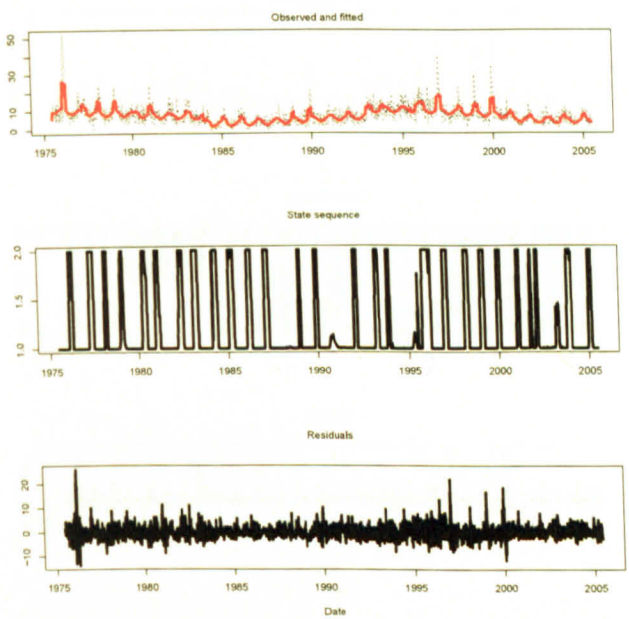


Figure 6.9.: **The state-sequence for the biIDR model fitted to P&I and laboratory reports in the 15-44 age group appears to converge and is clearly estimated.**

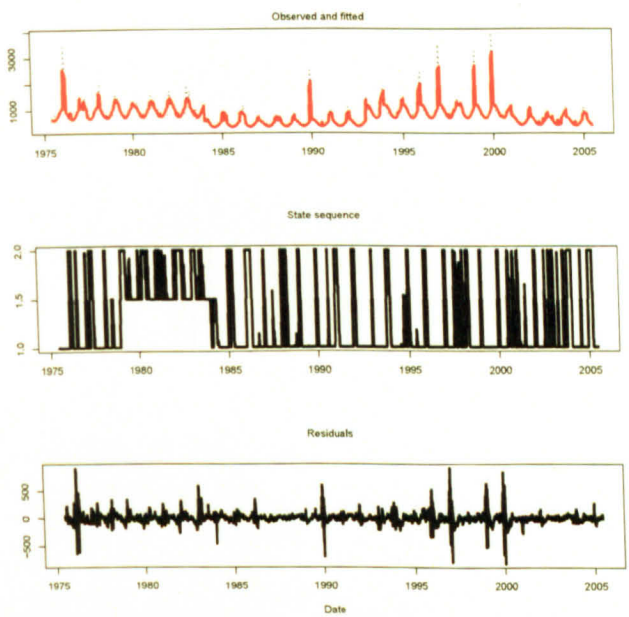


Figure 6.10.: **The state-sequence for the biLOGR model fit to P&I and laboratory reports from the ≥ 65 age group does not converge for the period between 1978/79 and 1982/83.**

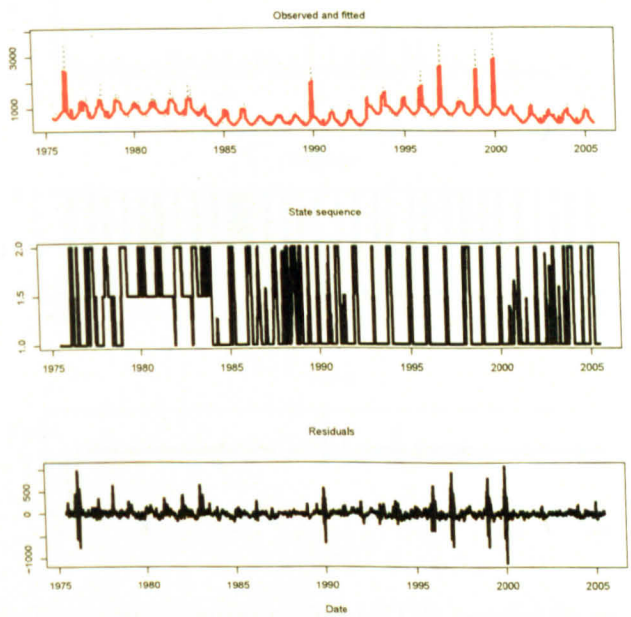


Figure 6.11.: **The state-sequence for the biIDR model fit to P&I and laboratory reports from the ≥ 65 age group (shown) does not converge for the period between 1978/79 and 1982/83, as for the BiLOGR fit.**

6.7.2. Models fitted to ILI and laboratory data

From fits to ILI data, state sequences from biLOGR fits appear to converge, as they do in the univariate case (e.g. figures E.6 and 5.18). State sequences from biIDR model fits to ILI from the 0-4, 5-14, 45-64 and $\geq$ 65 age groups do not converge (e.g. figure 6.12). The biIDR fit to ILI for the 15-44 age group is better than the univariate fit (figures 6.13 and 5.17).

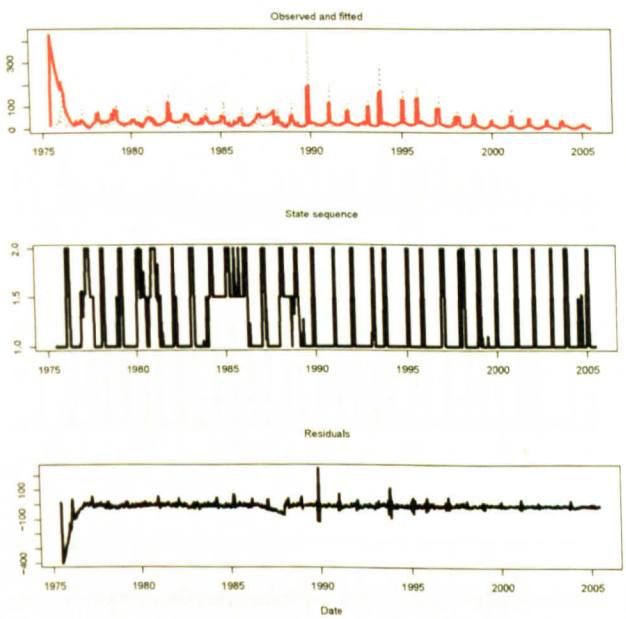


Figure 6.12.: **The state-sequence for biIDR model fit to ILI and laboratory reports in those 5-14 years of age does not converge for some weeks during several influenza seasons (1976, 1979, 1980, 1983, 1984, 1985, 1987, 1988 and 2004).**

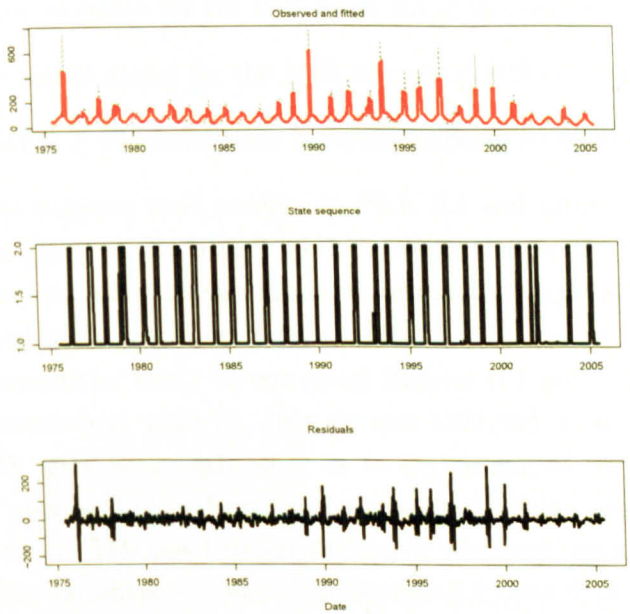


Figure 6.13.: **The state-sequence for biIDR model fit to ILI and laboratory reports for the 15-44 age group appears to converge and is clearly estimated.**

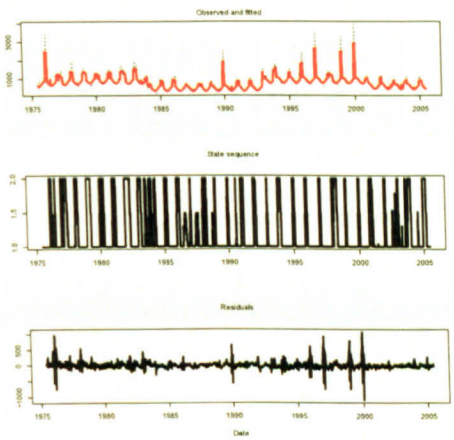
6.7.3. Models fitted to P&I, ILI and laboratory data

All triIDR state sequences appear to converge (e.g. figure 6.14). TriLOGR state sequences appear to converge (e.g. figure 6.15) with the exception of the ≥ 65 model which crashes after 30,000 iterations, having yet to converge (data not shown).

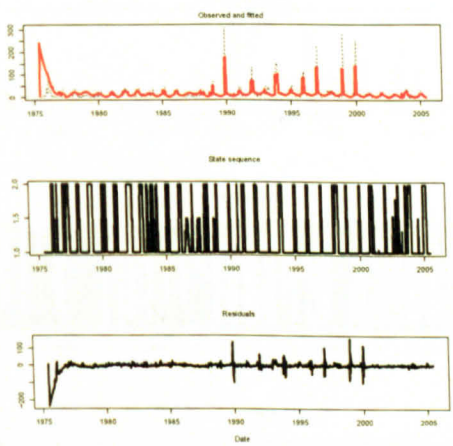
The crashing of the triLOGR ≥ 65 model may be related to the conflicts between P&I and laboratory data for the ≥ 65 age group in estimating a joint state sequence and the lack of convergence of the LOGR fit to laboratory reports for ≥ 65 mentioned previously. This conflict might also explain the following pattern of results from multivariate IDR fits to ≥ 65 data.

1. The state sequence for the IDR fit to P&I ≥ 65 appears to converge.
2. The state sequence for the IDR fit to ILI ≥ 65 does not converge.
3. Neither biIDR model for the ≥ 65 age group (P&I or ILI) converges.
4. The triIDR ≥ 65 model state sequence appears to converge.

This patterns suggests that pooling of P&I, ILI and laboratory data for the ≥ 65 age group in the triIDR model resolves the conflict between P&I and laboratory data for this age group in estimating a joint state sequence. (See overlayed state sequences for (univariate) IDR fits to P&I, ILI and laboratory reports for the ≥ 65 age group (figures G.1 and G.2). In several influenza seasons (e.g. 1990/91, 1996/97 and 1997/98), state sequences for P&I and ILI agree with each other as to the timing of the start and/or end of the 'aberrant' period and disagree with the state sequence for the laboratory data.) The lags between timing of 'aberrant' periods across the three outcome variables is revisited in sections 6.12 and 6.13.

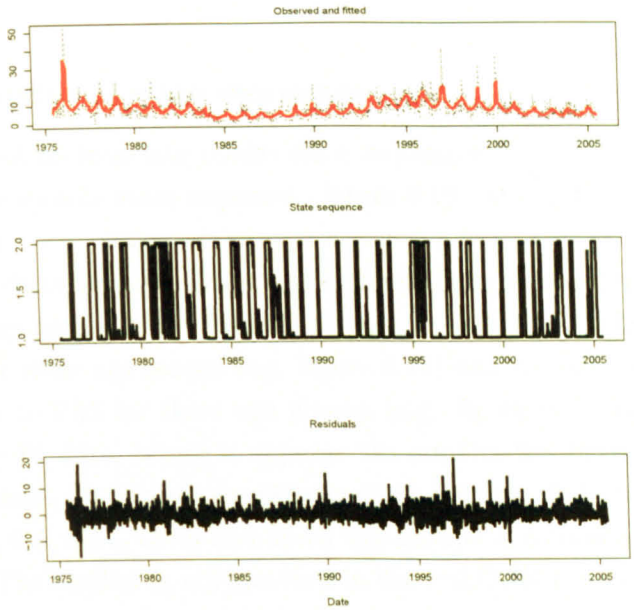


(a) P&I

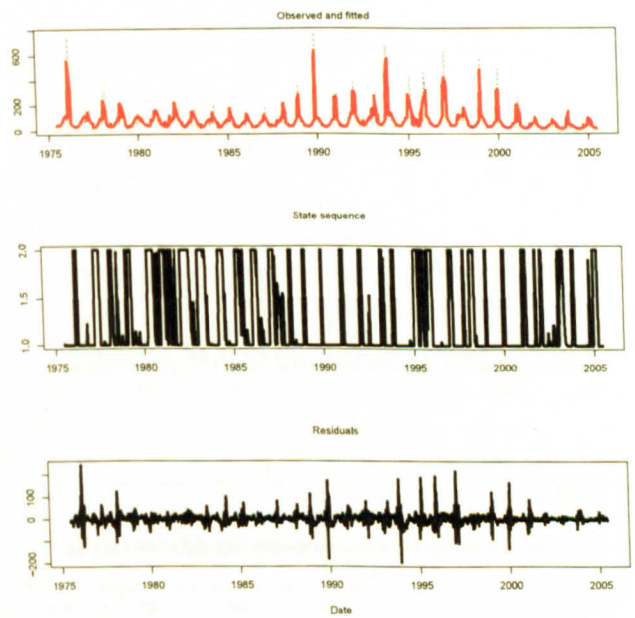


(b) ILI

Figure 6.14.: **The state-sequence from the triIDR model fit to data for the ≥ 65 age group appears to converge apart from a few weeks between the 1985/86 and 1986/87 seasons and between the 1986/87 and 1987/88 seasons.** The top plot shows triIDR model-predicted P&I for the ≥ 65 age group. The bottom plot shows triIDR model-predicted ILI for the ≥ 65 age group.



(a) P&I



(b) ILI

Figure 6.15.: **The state-sequence from the triLOGR model fit to data for the 15-44 age group appears to converge.** The top plot shows triLOGR model-predicted P&I for the 15-44 age group. The bottom plot shows triLOGR model-predicted ILI for the 15-44 age group.

6.7.4. Volatility of state sequences

Univariate and multivariate model state sequences vary in their volatility (e.g. of a less volatile state sequence - figure 6.16 - and a more volatile one - figure 5.14).

State sequences from triLOGR model fits to data for the 0-4, 5-14, 15-44 and 45-64 age groups (e.g. figure 6.17) are as volatile as biLOGR model fits to ILI for these age groups (e.g. figure 6.16) and are more volatile than biLOGR fits to P&I for these age groups (e.g. figure 6.7). In contrast to fits to the $\geq$ 65 data, where it appears the conflict lies between P&I and laboratory data, the above suggests conflict between ILI and laboratory data for the 0-4, 5-14, 15-44 and 45-64 age groups in estimating joint state sequences. The conflict is not resolved in the triLOGR models because the small numbers of deaths in the age groups under 65 contribute relatively little compared to the ILI and laboratory data.

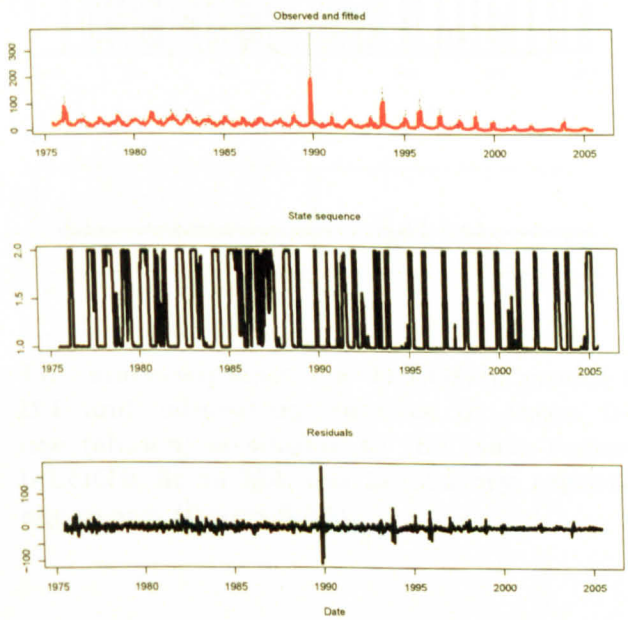


Figure 6.16.: **The state-sequence for biLOGR model fit to ILI and laboratory reports in those 0-4 years of age (shown) is more clearly estimated than LOGR fit (figure 5.14) and similar to the triLOGR model fit to data for the 0-4 age group (next figure).**

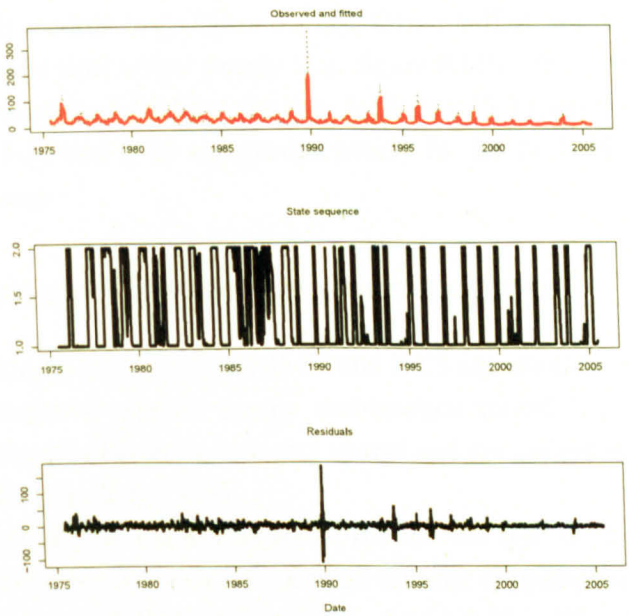


Figure 6.17.: **The state-sequence for triLOGR model fit to P&I, ILI and laboratory reports in those 0-4 years of age (shown) is similar to the state sequence for the biLOGR fit to ILI and laboratory reports in the 0-4 age group (figure 6.16).**

6.8. Plots of observed and fitted counts and residuals

Relative goodness of fit of models was determined by comparing plots of observed and fitted weekly counts and residuals. Plots were created by averaging results from the two chains.

biLOGR and triLOGR models predict the observed time series' of P&I and ILI well. Time series plots of residuals suggest there is underestimation of epidemics but that, otherwise, fit of biLOGR and triLOGR models is good (e.g. figure 6.10).

biIDR and triIDR models predict P&I from the 15-44 and $\geq$ 65 age groups and ILI from the 15-44 and 45-64 age groups as well as biLOGR and triLOGR models (e.g. figure 6.14a). Other biIDR and triIDR models predict observed time series' poorly (e.g. figure 6.14b). In particular, biIDR and triIDR models of P&I for the 0-4, 5-14 and 45-64 age groups and ILI for the 0-4, 5-14 and $\geq$ 65 age groups poorly fit the 1975/76, 1985/86 and 1987/88 seasons.

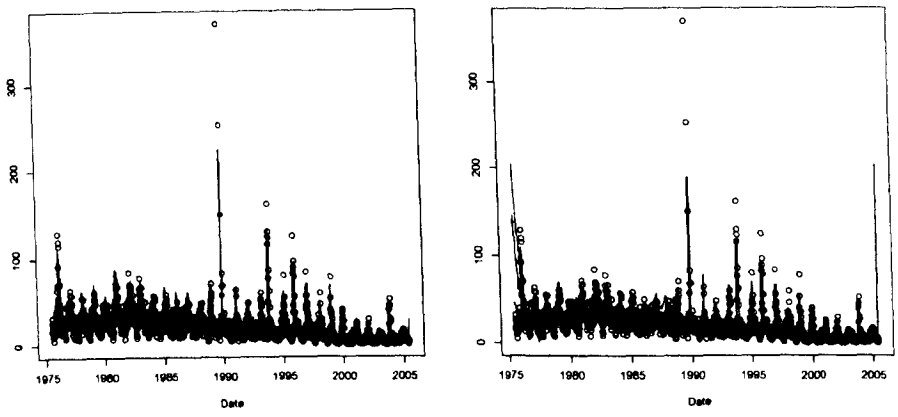
6.9. Posterior predictive density plots

Observed data for the 0-4, 15-44, 45-64 and $\geq$ 65 age groups are neither over- nor underdispersed relative to any multivariate model (e.g. figure 6.18). Deaths for the 5-14 age group are very sparse and are underdispersed relative to all models (e.g. figure E.8).

biIDR and triIDR models fitted to ILI for the 0-4, 5-14 and $\geq$ 65 age groups produce very wide CrIs for a small number of predicted counts during the 1975/76 and 2004/05 seasons (e.g. figure 6.18). This uncertainty in predicted counts is indicative of poor model fit of multivariate IDR models to ILI for these age groups noted in the previous section.

6.10. Autocorrelation plots of residuals

No model accounts for all autocorrelation in any of the modeled data sets. Multivariate LOGR model fits are generally adequate, leaving some positive residual autocorrelation at lags of approximately 1-2 weeks, because of underfitting epidemics, and at 52 ± 2 and 104 ± 2 weeks because of failure

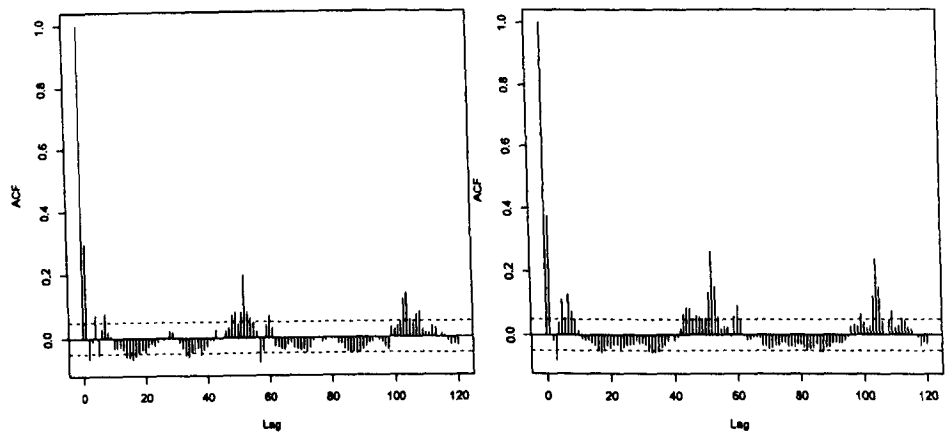


(a) triLOGR (b) triIDR

Figure 6.18.: **Posterior predictive density plots of triLOGR and triIDR models fitted to ILI from the 0-4 age group.** The observed data do not all fall within posterior predictive CrIs (meaning ILI data for the 0-4 age group are not underdispersed relative to the trivariate models). Neither do the majority of the data lie outside posterior predicted CrIs (meaning ILI data for the 0-4 age group are not overdispersed relative to the trivariate models).

to model seasonal variation sufficiently (e.g. figure 6.19). The triLOGR model fit to data for the $\geq$ 65 age group is poor, with residual autocorrelation at most lags up to approximately 52 weeks and at 100-105 weeks (e.g. figure 6.20).

Multivariate IDR model fits to P&I are similar to multivariate LOGR fits, in terms of residual autocorrelation (e.g. figure 6.19). The triIDR model better fits P&I data for the $\geq$ 65 age group than the triLOGR model (figure 6.20). Multivariate IDR models fitted ILI for the 0-4, 5-14 and $\geq$ 65 age groups poorly, leaving autocorrelation in residuals to lags of up to 45 weeks (e.g. figure E.9). This corroborates poor model fit to these data noted in the previous two sections.

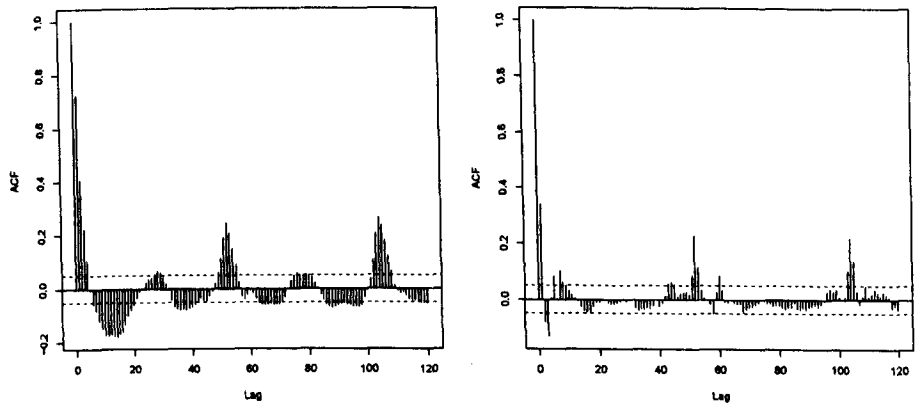


(a) biLOGR (b) biIDR

Figure 6.19.: **Autocorrelation plots of residuals from biLOGR and biIDR model fits to P&I and laboratory reports from the $\geq$ 65 age group.** Adequate model fit of the data is evident apart from underestimation of large peaks (shown by the positive autocorrelation at lag 1 week) and insufficient modeling of seasonality (shown by positive autocorrelation at lags around 52 and 104 weeks).

6.11. Precision of the mean shift random effect

The means of the posterior distributions for the precisions of the mean shift random effects, τ_{rand}, for fits to P&I and ILI were compared between univariate, bivariate and trivariate models (tables 6.4 to 6.7). This was done



(a) triLOGR (b) triIDR

Figure 6.20.: **Autocorrelation plots of residuals from triLOGR and triIDR model fits to P&I data for the $\geq$ 65 age group.** The triIDR fit (b) is adequate (with positive autocorrelation at lags around 1, 52 and 104 weeks only). The triLOGR fit (a) is poor (with positive and negative autocorrelation at many lags).

to determine whether modeling more than one outcome variable in a joint model resulted in increased power to estimate the mean shifts because of sharing of information across outcome variables about state transitions.

Table 6.4.: **Mean of the posterior distribution for the precision of the mean shift random effect (τ_{rand}): LOGR models fitted to P&I.** Mean posterior precision increases from LOGR to biLOGR fits for the 0-4 and 15-44 age groups. Blanks indicate the posterior precision parameter does not converge.

	Mean of posterior for τ_{rand}	**SD**	**MC-error**	**2.5% credible bound**	**Median of posterior for τ_{rand}**	**97.5% credible bound**
0-4, LOGR	175.00	250.70	7.38	22.68	105.50	790.50
0-4, biLOGR	353.10	349.70	22.60	48.39	235.90	1293.00
0-4, triLOGR	172.00	145.30	7.53	36.65	128.00	574.50
5-14, LOGR	397.80	503.30	12.68	20.17	213.60	1771.00
5-14, biLOGR	261.30	415.80	16.95	25.79	119.20	1497.00
5-14, triLOGR	216.10	405.90	23.99	15.91	83.26	1203.00
15-44, LOGR	21.84	9.64	0.17	9.28	19.93	45.18
15-44, biLOGR	32.68	10.14	0.49	17.14	30.73	56.85
15-44, triLOGR	22.77	7.22	0.21	11.22	22.19	40.00
45-64, LOGR						
45-64, biLOGR	16.05	4.59	0.22	8.31	15.62	26.13
45-64, triLOGR	11.46	2.96	0.15	6.20	11.01	17.82
≥ 65, LOGR						
≥ 65, biLOGR						
≥ 65, triLOGR	7.12	5.95	0.82	0.86	3.89	18.57

Table 6.5.: **Mean of the posterior distribution for the precision of the mean shift random effect (τ_{rand}): LOGR models fitted to ILI.** Precision increases from the LOGR to biLOGR fits for the 0-4, 15-44 and 45-64 age groups.

	Mean of posterior for τ_{rand}	**SD**	**MC-error**	**2.5% credible bound**	**Median of posterior for τ_{rand}**	**97.5% credible bound**
0-4, LOGR	3.91	1.04	0.01	2.14	3.81	6.19
0-4, biLOGR	4.24	1.15	0.03	2.35	4.08	6.73
0-4, triLOGR	4.39	1.26	0.06	2.38	4.21	7.23
5-14, LOGR	6.10	1.65	0.02	3.31	5.96	9.71
5-14, biLOGR	6.06	1.58	0.05	3.39	5.93	9.62
5-14, triLOGR	6.09	1.69	0.08	3.27	6.00	9.63
15-44, LOGR	5.48	1.75	0.09	2.77	5.24	9.52
15-44, biLOGR	6.30	1.58	0.13	3.47	6.06	9.41
15-44, triLOGR	6.26	1.71	0.06	3.53	6.06	9.90
45-64, LOGR	5.64	1.50	0.02	3.12	5.51	8.92
45-64, biLOGR	6.23	1.73	0.05	3.45	5.99	9.97
45-64, triLOGR	6.09	1.59	0.07	3.46	5.99	9.56
≥ 65, LOGR	4.26	1.16	0.02	2.31	4.15	6.79
≥ 65, biLOGR	3.35	0.95	0.04	1.83	3.25	5.83
≥ 65, triLOGR	1.24	1.24	0.18	0.05	0.92	3.55

Table 6.6.: **Mean of the posterior distribution for the precision of the mean shift random effect (τ_{rand}): IDR models fitted to P&I.** Precision increases from IDR to biIDR fits for all age groups.

	Mean of posterior for τ_{rand}	SD	MC-error	2.5% credible bound	Median of posterior for τ_{rand}	97.5% credible bound
0-4, IDR	2.81E-01	1.61E-01	5.12E-03	1.02E-01	2.42E-01	6.84E-01
0-4, biIDR	4.89E-01	2.06E-01	8.94E-03	1.96E-01	4.58E-01	1.02E+00
0-4, triIDR	7.74E-01	4.55E-01	3.87E-02	2.70E-01	6.78E-01	2.04E+00
5-14, IDR	1.20E+02	3.18E+02	1.49E+01	3.28E-01	1.16E+01	1.01E+03
5-14, biIDR	4.98E+01	1.70E+02	1.04E+01	3.42E+00	1.41E+01	3.31E+02
5-14, triIDR	2.31E+02	3.60E+02	2.26E+01	1.34E+01	9.59E+01	1.36E+03
15-44, IDR	3.79E-02	1.64E-02	3.25E-04	1.48E-02	3.51E-02	7.65E-02
15-44, biIDR	1.17E-01	4.03E-02	2.02E-03	5.64E-02	1.11E-01	2.09E-01
15-44, triIDR	6.39E-02	2.06E-02	7.06E-04	3.30E-02	6.05E-02	1.07E-01
45-64, IDR	1.12E-03	3.28E-04	6.26E-06	5.78E-04	1.08E-03	1.86E-03
45-64, biIDR	1.86E-03	5.44E-04	1.74E-05	1.01E-03	1.80E-03	3.11E-03
45-64, triIDR	1.41E-03	3.81E-04	1.51E-05	7.58E-04	1.37E-03	2.23E-03
≥ 65, IDR	1.86E-06	4.83E-07	5.10E-09	1.02E-06	1.82E-06	2.94E-06
≥ 65, biIDR	1.94E-06	5.16E-07	2.94E-08	1.07E-06	1.88E-06	3.03E-06
≥ 65, triIDR	1.98E-06	4.97E-07	2.84E-08	1.06E-06	1.94E-06	3.06E-06

Table 6.7.: **Mean of the posterior distribution for the precision of the mean shift random effect (τ_{rand}): IDR models fitted to ILI.** Precision increases from IDR to biIDR fits for all age groups.

	Mean of posterior for τ_{rand}	**SD**	**MC-error**	**2.5% credible bound**	**Median of posterior for τ_{rand}**	**97.5% credible bound**
0-4, IDR	5.57E-04	1.57E-04	2.01E-06	2.95E-04	5.42E-04	9.10E-04
0-4, biIDR	1.20E-03	3.29E-04	1.68E-05	6.63E-04	1.17E-03	1.99E-03
0-4, triIDR	1.29E-03	3.52E-04	1.37E-05	6.64E-04	1.27E-03	2.04E-03
5-14, IDR	3.81E-04	1.19E-04	2.49E-06	1.91E-04	3.67E-04	6.57E-04
5-14, biIDR	5.17E-04	1.53E-04	4.86E-06	2.70E-04	5.02E-04	8.56E-04
5-14, triIDR	5.67E-04	1.69E-04	1.06E-05	2.88E-04	5.47E-04	9.30E-04
15-44, IDR	3.11E-05	8.40E-06	1.11E-07	1.72E-05	3.02E-05	4.95E-05
15-44, biIDR	3.16E-05	9.04E-06	4.15E-07	1.73E-05	3.05E-05	5.37E-05
15-44, triIDR	3.21E-05	8.57E-06	3.05E-07	1.78E-05	3.12E-05	5.11E-05
45-64, IDR	1.12E-04	3.48E-05	1.30E-06	5.66E-05	1.07E-04	1.91E-04
45-64, biIDR	1.76E-04	5.36E-05	1.74E-06	8.60E-05	1.71E-04	2.94E-04
45-64, triIDR	1.85E-04	5.38E-05	1.85E-06	9.26E-05	1.79E-04	3.00E-04
$\geq$ 65, IDR	4.25E-04	1.30E-04	3.05E-06	2.08E-04	4.13E-04	7.20E-04
$\geq$ 65, biIDR	6.28E-04	1.83E-04	5.72E-06	3.33E-04	6.08E-04	1.03E-03
$\geq$ 65, triIDR	4.98E-04	1.36E-04	8.75E-06	2.76E-04	4.84E-04	8.13E-04

The mean posterior precision of the mean shift random effect generally increases from LOGR to biLOGR and from IDR to biIDR models. There is considerable uncertainty in estimates of posterior precision across models. There is no increase in mean shift random effect precision for biLOGR model fits to P&I and ILI data for the 5-14 age group and ILI data for the $\geq$ 65 age group relative to LOGR fits. Deaths and laboratory reports from the 5-14 age group have many zeros making these data difficult to model. BiLOGR and triLOGR fits to the $\geq$ 65 data are problematic as discussed in sections 6.7 and 6.10.

TriLOGR fits to most age groups do not show an additional increase in posterior precision of the mean shift random effect over the biLOGR models, probably because triLOGR model fits are generally the same as, or poorer than, biLOGR fits (tables 6.4 and 6.5, section 6.7.4). For triIDR model fits to most age groups there is an additional increase in the posterior precision of the mean shift random effect, compared with biIDR model fits (tables 6.6 and 6.7).

6.12. Exploration of lagged state transitions between outcomes

To determine whether there is evidence that, for a given influenza season in a given age group, state transitions for one outcome variable are lagged relative to state transitions for the other outcome variables, state sequences for univariate fits to each age group for P&I, ILI and laboratory reports were overlayed, separately for LOGR and IDR model fits. This was done to allow that any lag might vary between influenza seasons.

State transitions in a given influenza season do not happen in the same week across outcome variables within an age group (e.g. figure 6.21). This is due, for example, to the lag between seeking care for an influenza illness and dying from one (section 3.2.6). The relative timing of transitions across age groups is not consistent across influenza seasons. Biological explanations for the inconsistent lag between outcome variables across influenza seasons may relate to the dominant circulating viruses in different influenza seasons and to the relative virulence of circulating viruses in different seasons. Artifactual explanations for the inconsistent lag between outcome variables

across influenza seasons may include changes to the proportion of laboratory reports from GPs over time which resulted in a change to the profile of patients represented in laboratory data over time in terms of, for example, comorbidities.

State sequences are more clearly estimated from bivariate model fits than from univariate model fits, so bivariate state sequences were also overlayed to check for evidence of lagged state transitions (separately for biIDR and biLOGR models). State transitions do not happen in the same week for P&I and ILI in a particular age group and different lags are observed each season (e.g. figure 6.22).

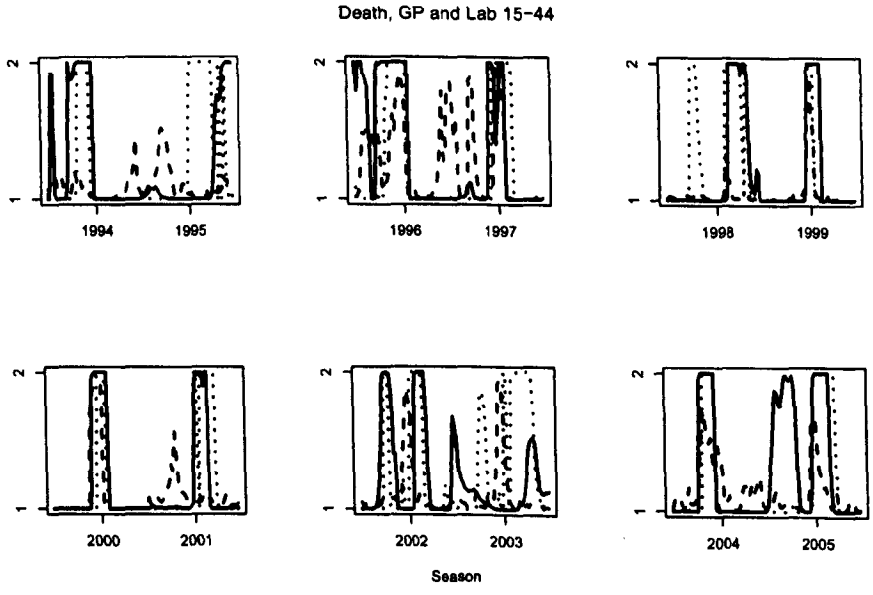


Figure 6.21.: **State sequence from LOGR fits to P&I (dashed line), ILI (dotted line) and lab reports for influenza A (solid line) from the 15-44 age group: 1993/04-2004/05.** The state sequences do not converge for all seasons. There is a lag between state transitions for P&I, ILI and lab reports and the lag varies by influenza season.

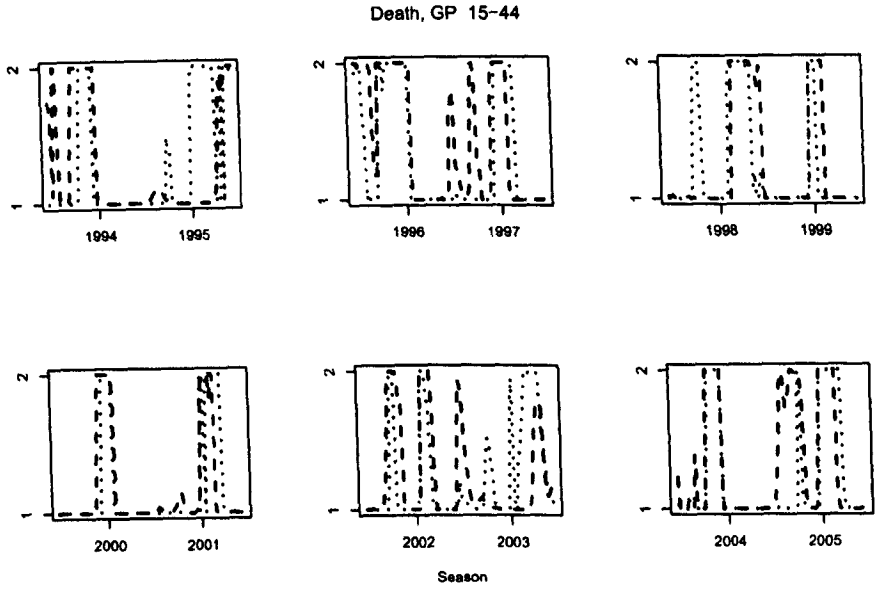


Figure 6.22.: **State sequences from biLOGR fits to P&I (dashed line) and ILI (dotted line) from the 15-44 age group: 1993/04-2004/05.** There is a lag between state transitions for ILI and P&I and the lag varies by influenza season.

6.13. Models with a lag

Given evidence for a lag between state transitions for different outcome variables, multivariate LOGR and IDR models allowing state transitions for different outcome variables to be lagged relative to one another by a constant $\pm$ 1 or 2 weeks for the duration of the time series were explored (the model formula is shown in Appendix G). In bivariate models, the average lag between state transitions in P&I or ILI relative to state transitions in laboratory reports was estimated from the data. In trivariate models, two lags - the average lag between state transitions in P&I relative to state transitions in laboratory reports and the average lag between state transitions in ILI and in laboratory reports - were estimated from the data. For every model, the 5 possible lags ($\pm$ 2, 1 or 0 weeks) were assigned equal prior probability ($p = 0.20$). Models allowing for a lag take 2 orders of magnitude longer to complete an iteration of the sampler compared with models without a lag. Models were run for 10,000 iterations on two chains, saving 10% of sampled values, to determine if such models would run in OpenBUGS given the added complexity. Models have yet to converge after 10,000 iterations as expected since models without a lag take longer than 10,000 iterations to converge (e.g. figure 6.23). Models including a lag were not taken forward for estimating the impact of covariates in the next chapter because evidence presented in section 6.12 suggests an inconsistent lag from season to season. An inconsistent lag from influenza season to influenza season is not straightforward to model in this framework; this was not attempted.

6.14. Summary of results

In general, multivariate LOGR models are better than multivariate IDR models; most of the biLOGR and triLOGR models appear to converge, as opposed to many of the multivariate IDR models. Multivariate LOGR models also fit the data better than multivariate IDR models. Modeling P&I or ILI jointly with laboratory reports increases the precision of the random effect mean shift for most model fits. In triIDR models there appears to be additional sharing of information across outcomes, relative to the biIDR models, for estimating state sequences and thus mean shifts. There is much

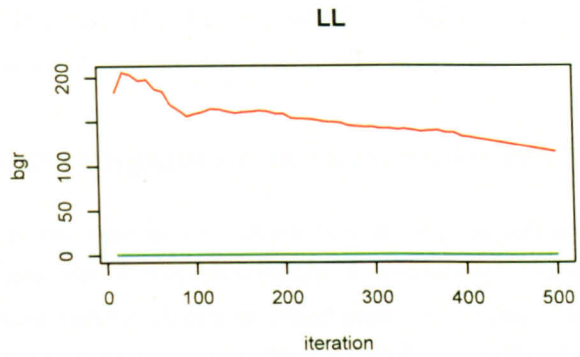


Figure 6.23.: **BGR plot of the LL of triLOGR model with lag, fitted to the 0-4 age group, after 10,000 iterations.** The ratio of pooled to within chains variability (the top (red) line) has yet to converge, i.e. come to 1, after 10,000 iterations. The plot is based on a sample of every 10th iteration from the last 5000 iterations of the run.

uncertainty in the posterior distribution for the mean shift random effect. Because of problems with convergence and fit, however, no multivariate IDR models were taken forward for estimating covariate effects in chapter 7. In triLOGR models, precision of the mean shift declines relative to biLOGR fits. There appears to be a conflict between ILI and laboratory reports for age groups <65 and between P&I and laboratory reports for the $\geq$ 65 age group in jointly estimating state sequences. This is most likely because the timing of 'aberrant' periods differs across outcome variables for a given age group. Two observations support this explanation. First, overlaying univariate state sequences shows state transitions occur during different weeks across the outcome variables and that the lag varies from one influenza season to the next. Second, biLOGR and triLOGR state sequences are generally more volatile than univariate state sequences. This is consistent with there being conflict between outcome variables in estimating the state sequence due to differences in the timing of 'aberrant' periods in different outcome variables. OpenBUGS accommodates the increased complexity of fitting models allowing a constant lag between state transitions across

outcomes. A longer time series (i.e. more than the 30 influenza seasons analyzed) would probably be required to estimate a state sequence where a time-varying lag was allowed between state transitions across outcome variables, though proof of concept could be shown using simulated data.

BiLOGR and triLOGR models were taken forward for estimating impact of CT seasons and vaccination on the mean shift in chapter 7.

6.15. Strengths of multivariate HMMs

Multivariate models allowed sharing of information across outcome variables for estimation of state sequences, as evidenced by apparent convergence of more multivariate than univariate state sequences. This is especially obvious comparing state sequences for biLOGR to LOGR fits for P&I from the 0-4 and 5-14 age groups. LOGR models are unable to distinguish 'aberrant' from 'normal' P&I incidence for the 0-4 or 5-14 age groups. BiLOGR models are clearly able to do so. Bivariate models increase precision of mean shifts relative to univariate models. Multivariate HMMs increase specificity of P&I and ILI for influenza.

6.16. Limitations of multivariate HMMs

There is a time-varying lag between state transitions for outcome variables within an age group and this is not allowed for in the models. As in the previous chapter, inadequate modeling of large peaks in incidence and of seasonality is obvious in autocorrelation plots of residuals.

7. Estimating the effect of cluster transitions and the impact of vaccination on P&I and ILI using joint HMMs

7.1. Aims of this chapter

The aim of the work described in this chapter was to use biLOGR and triLOGR HMMs developed in chapter 6 to quantify the effect of CT seasons in the antigenic evolution of influenza A/H3N2 virus, and the impact of rising vaccination coverage of the elderly in England & Wales, on the mean shift in P&I and ILI across age groups.

7.1.1. Objectives of this chapter

1. To fit biLOGR and triLOGR HMMs to each age group including dependency between the random effect mean shift and
   a) CT seasons as a binary variable, or
   b) antigenic distance between CT seasons as a quantitative variable, or
   c) vaccine coverage in the ≥ 65 age group each influenza season as a quantitative variable
2. To carry out sensitivity analysis on priors used for covariate effects in the models above
3. To discuss confounding and effect modification of the effect of CT seasons on mean shifts and of the impact of vaccination on mean shifts

7.1.2. Main findings

Scatter plots of crude mean shifts against exposures of interest - CT seasons, the antigenic distance between clusters and vaccine coverage of the elderly - suggest little evidence of association between exposures of interest and mean shifts in P&I or ILI. This agrees with the finding of little crude association between these exposures of interest and peak incidence in P&I or ILI noted in chapter 4. It was not possible to quantify the magnitude of exposure effects of interest using the biLOGR and triLOGR HMMs. There is limited information in the data about effects of interest. Posterior distributions for coefficients for vaccine impact and CT effect on mean shifts are little influenced by the data and more influenced by priors.

7.2. Introduction

Peak P&I and ILI incidence observed in a given influenza season is highly variable. Several factors may affect peak P&I and ILI incidence, including CT seasons [14, 94] and increasing vaccination coverage of the elderly. [7] In chapter 4 it was noted that, crudely, the distribution of peak rates for the first H3N2-dominated season after a CT appears greater than for intracluster seasons. Ranking of peak P&I and ILI rates observed across seasons revealed that only for P&I in the $\geq$ 65 age group do CT seasons occur in at least five of the top ten seasons. T-tests suggest weak evidence for small increases in peak P&I and ILI in the first H3N2-dominated season after a CT compared with the average season (6 more P&I per 1,000,000 population and 95 more ILI per 100,000 population) and compared with intracluster seasons (8 more P&I per 1,000,000 and 133 more ILI per 100,000). These small differences are of little public health importance given that, in the data analyzed in the thesis (excluding the 1969/70 pandemic season), weekly rates of P&I of up to 80/1,000,000 and ILI of up to 2,322/100,000 are observed. It was also noted in chapter 4 that there is no clear association between peak rates of P&I and ILI each influenza season and the antigenic distance between clusters. There is a weak negative association between vaccine coverage of the $\geq$ 65 age group and peak seasonal P&I and ILI rates across age groups.

7.3. Covariates

Seasons when new antigenic clusters of influenza A/H3N2 viruses become dominant may plausibly experience larger average mean shifts than intra-cluster seasons because of the increase in the proportion of the population susceptible to H3N2. Each unit of antigenic distance within a CT may also have a detectable effect on the mean shift. Increasing vaccine coverage of the ≥ 65 age group would be expected to lead to smaller mean shifts in the ≥ 65 age group, and plausibly in the other age groups, though relatively low vaccine effectiveness [113] and the fact that it is likely to be children, not the elderly, who are drivers of transmission in the community [38] suggest any impact of vaccination on other age groups would be approximately nil. If all influenza - attributable incidence is not designated as being from the 'aberrant' state there may be residual influenza - attributable incidence within the 'normal' state. If this is so, an effect of CT seasons/antigenic distance and an impact of vaccination might also be distinguishable in changes in 'normal' incidence. Another consequence of misclassification of influenza - attributable incidence to the 'normal' state, if present, would be lower power to detect exposure effects on the mean shift. As a first step in exploring whether CT seasons/antigenic distance or vaccine coverage of the elderly explains some of the variability in the impact of influenza seasons in terms of morbidity or mortality, the random effect mean shift was expressed as dependent on each of CT seasons, the antigenic distance between clusters and vaccine coverage of the ≥ 65, in separate model runs. CT seasons were defined as the first H3N2-dominated season after a CT. The effect of antigenic distance between clusters on the mean shift was explored to allow for the fact that CT seasons differ in size (range 3.3 to 7.8 antigenic units). [13]

7.4. Description of the model

BiLOGR and triLOGR models were fitted to P&I and ILI data using models identical to chapter 6 (in terms of structure and priors) apart from the dependency between the mean shift and exposures of interest. Exposure effects on the mean shift were modeled by expressing the mean of the random effect as dependent on the exposures like this

$$
\begin{aligned}
&\text{for(i in 1:30)\{} \\
&\qquad \alpha_1[\text{i}] \sim \text{Normal}(\mu_{\text{rand}}[\text{i}], \tau_{\text{rand}})\text{I}(0,) \\
&\qquad \mu_{\text{rand}}[\text{i}] = \mu_{\text{rand}_0} + \mu_{\text{rand}_1} * \text{CT}[\text{i}] \\
&\qquad \} \\
&\qquad \mu_{\text{rand}_0} \sim \text{Normal}(0.0, 1E-2) \\
&\qquad \mu_{\text{rand}_1} \sim \text{Normal}(0.0, 1E-3) \\
&\qquad \tau_{\text{rand}} \sim \text{Gamma}(0.001, 0.001)
\end{aligned}
$$

where i denotes an influenza season and CT is a binary variable coding influenza seasons as 1 for the first H3N2-dominated season after a CT and 0 for intracluster seasons.

Coefficients for exposure effects are the average effect of, for example, a CT on the average mean shift. Exponentiated (exp) regression coefficients for covariates are ratios of the average mean shift with to without the covariate (or per unit change in the covariate). Estimating the effect of covariates on the mean shift using the multivariate LOGR model framework implies that the association between the exposure and the mean shift is additive on the logarithmic scale.

7.5. Prior sensitivity analysis

A prior sensitivity analysis was carried out to test whether the association between exposures of interest and mean shifts were robust to the choice of priors on coefficients for exposures. First, covariate effects were given reference Normal(0.0, 1E-3) priors. Second, weakly informative priors were set using estimates of the variability in excess P&I and ILI from other settings (sections 7.5.1 and 7.5.2). Finally, strongly informative priors on the effect of CTs, and the antigenic distance between clusters, on the mean shift were set based on the results from a previous model fitted by Koelle *et al.* (section 7.5.3). [14]

All weakly informative priors were uniform distributions. Uniform distributions assign approximately equal probability to all values within a range. CT seasons plausibly result in an instantaneous increase in the proportion of the population susceptible to circulating influenza A/H3N2 viruses and so would be expected to result in larger mean shift than seasons not experiencing a CT. Descriptive studies of the impact of large antigenic drift events

suggest they at least sometimes coincide with epidemics (e.g. [93]). Each unit change in the antigenic distance between clusters could also result in a detectable increase in the size of the susceptible population and thus an inflated mean shift. Higher vaccine coverage should lead to dampening of the mean shift.

The aim of the analysis in this chapter was to determine whether the exposures of interest explain some of the variability in the impact of influenza seasons, as captured by the mean shift. The magnitude of exposure effects of interest should lie within the maximum rate ratio (MAXRR) of excess ILI or P&I observed between two influenza seasons in a single setting. Estimates of MAXRR (based on different data from those analyzed for the thesis) were used to set weakly informative priors. MAXRR was used to establish an upper bound for the possible effect of CT seasons (or a unit change in antigenic distance between clusters) on the mean shift. The reciprocal of MAXRR was used to establish a lower bound on vaccine impact. Using MAXRR to set extremes of possible covariate effects assumes no negative confounding of the covariate effect.

A literature search was done to identify reports of excess inter-pandemic ILI or P&I from other temperate Northern hemisphere settings with which to estimate MAXRR. A search of Pubmed on June 9th 2009 using the terms "excess", "influenza - attributable", "influenza-associated", "influenza epidemic" (and variations on these terms, separated by "OR") with "influenza" identified 13 papers based in a temperate Northern hemisphere setting, or settings, that provided estimates of excess P&I or ILI counts or rates in two or more influenza seasons after 1969/70 (the pandemic season). [7,8,54,56, 58,61,65,73,92,107,197–199] All of these studies reported estimates of excess P&I counts or rates for at least two influenza seasons. One study reported estimates of excess ILI for at least two influenza seasons. [65] The inclusion criterion requiring that studies report on at least two influenza seasons was required so that study-specific MAXRR estimates could be derived. Published excess counts were translated into equivalent rates per 100,000 for ILI, and per 1,000,000 for P&I, for England & Wales using population estimates from national statistics websites. Where studies reported an excess of zero, 1 excess death was added to these seasons in order to be able to calculate MAXRR (dividing by zero is undefined).

MAXRR for P&I from studies reporting on all-ages or on the $\geq$ 65 age

group was 24,800 (122/1,000,000 in 1972/73 divided by 0.004/1,000,000 in 1970/71 and 1973/74). [198] MAXRR for ILI was 785,252 (1,387/100,000 per week in 1989/90 divided by 0.002/100,000 per week in 1999/00). [65]

7.5.1. Weakly informative priors on the effect of CTs on the mean shift

As mentioned above, it is plausible that CTs result in an instantaneous rise in the proportion of the population susceptible to circulating influenza A/H3N2 viruses, thus inflating the mean shift. It is less plausible that CT seasons result in smaller mean shifts on average.

The weakly informative prior for the coefficient of the effect of CT seasons, or a unit change in antigenic distance between clusters, on the mean shift was set as a uniform prior with the log(MAXRR) as the upper bound. The lower bound was arbitrarily set as the 10th root of the upper bound, in the opposite direction. This was done in order not to constrain the prior such that CTs could only cause an increase (or no change) in the average mean shift. In this way the prior was 'weakly informative' of the effect of CTs on the mean shift and allowed a small probability that CTs cause a dampening of the mean shift. On the logarithmic scale, this translated to a weakly informative prior on the average effect of cluster transitions, or per unit change in antigenic distance between clusters, on the mean shift in P&I of $\sim$ Uniform$(-1.01, 10.12)$ and in ILI of $\sim$ Uniform$(-1.36, 13.57)$.

7.5.2. Weakly informative priors on vaccine impact

The impact of increasing vaccine coverage of the elderly on the mean shift was first explored with vaccine coverage as a quantitative variable. As mentioned in section 7.5, the dampening impact on the mean shift per unit increase in vaccine coverage should be no more extreme than log(1/MAXRR). Models with a weakly informative upper bound on the prior - for example, which restricted a unit increase in vaccine coverage to cause an inflation of the mean shift with only a small probability - were computationally problematic (see below). The upper bound of the weakly informative prior on the impact of each unit increase in vaccine coverage on the mean shift was therefore set at log(MAXRR). On the logarithmic scale, this translated to a weakly informative prior on the average effect of a unit change

in vaccine coverage of the $\geq$ 65 age group on the mean shift in P&I of $\sim$ Uniform$(-10.12, 10.12)$ and in ILI of $\sim$ Uniform$(-13.57, 13.57)$.

The reason why it was not possible to set a weakly informative upper bound on the prior for the impact of vaccine coverage, as a quantitative variable, on the mean shift may be that the assumption of an additive (on the logarithmic scale) impact of each unit increase in vaccine coverage on the mean shift is invalid. Modeling vaccine coverage as an ordered categorical, instead of quantitative, variable eased computational difficulties. There are natural step changes in the vaccine coverage data that were used to define levels of the categorical variable. The first step change occurred in 1989/90: before 1989/90, vaccine coverage of the $\geq$ 65 age group was assumed to be zero. Between 1989/90 and 1999/00, yearly vaccine coverage of the $\geq$ 65 age group ranged from 25% (in 1989/90) to 47% (in 1999/00). The second step change in coverage occurred in 2000/01. From 2000/01, yearly vaccine coverage of the $\geq$ 65 age group ranged from 67% (in 2000/01) to 73% (in 2003/04). By fitting models with a dependency between the mean shift and vaccine coverage of the elderly as a ordered categorical variable, the average mean shift in seasons between 1975/76 and 1988/89 (the reference period when coverage was assumed to be zero) was compared to the average mean shift in seasons with moderate vaccine coverage of the elderly (1989/90 to 1999/00) as well as to the average mean shift in seasons with high vaccine coverage of the elderly (2000/01 to 2004/05). These models were fitted with both a reference prior and a weakly informative prior on the coefficients for each step change in vaccine coverage on the mean shift in separate model fits. The reference prior was the same prior used in the model with vaccine coverage as a quantitative variable ($\sim$ Normal$(0.0, 1E-3)$). The weakly informative prior was $\sim$ Uniform$(-13.57, 1.36)$. Note that in this case it was possible to constrain the upper bound of the weakly informative prior to allow little prior probability that increasing vaccine coverage leads to larger mean shifts.

7.5.3. Strongly informative priors on the effect of CTs on the mean shift

The strongly informative prior on CT seasons was the natural log of a prior which, on the original scale, has median 1.66 (the ratio of peak height in CT

vs. intracluster seasons from [14]) and 95% prior mass between 0.8 and 4. The range of this prior was set to encompass a plausible range of estimates of the effect of CT seasons without being too informative. On the original scale, this translated to a strong prior on the average effect of CTs on the mean shift in both P&I and ILI of $\sim$ Normal$(1.6625, 0.7)$I$(0.7,)$.

The strongly informative prior on the antigenic distance between clusters was the log of a prior which, on the original scale, had median 1.11 (calculated as the median of the CT effect prior, 1.66, to the power of 1/4.5 ($1.66^{1/4.5}$), where 4.5 is number of antigenic units in the average CT [13]) and 95% prior mass between 0.8 and $1.66^{1/3.3}$ (where 3.3 is the smallest CT in [13]). On the original scale, this translated to a strong prior on the average effect per unit change in the antigenic distance between clusters on the mean shift in both P&I and ILI of $\sim$ Normal$(1.119588, 10)$I$(0.7, 1.53)$.

7.5.4. Strongly informative priors on vaccine impact

Strongly informative priors on the impact of vaccination were not explored in models of the impact per unit increase in coverage because of the difficulties in setting weakly informative priors (section 7.5.2). Strongly informative priors on the impact of moderate or high coverage as an ordered categorical variable were not explored because results from reference and weakly informative priors clearly indicate there is insufficient information in the analysis of mean shifts in ILI for the $\geq$ 65 age group to be able to quantify vaccine impact. It was not possible to estimate covariate effects on P&I for the $\geq$ 65 age group because multivariate LOGR models fitted to these data do not converge (tables 6.2 and 6.3). Any impact on other age groups of vaccination of the elderly would be expected to be much smaller than the impact in the elderly and therefore more difficult to detect.

7.6. Model convergence

Two chains for each parameter were started from different sets of initial values and were run for 100,000 iterations, saving 1% of sampled values. Initial values were the same as those used to initialise models in chapter 6. Model fit was assessed based on plots of the final 500 saved samples. Coefficients for the three exposures of interest (CT seasons, the antigenic distance between

clusters and vaccine coverage of the those $\geq$ 65) were only interpreted for models that appear to converge and to clearly estimate the state sequence (i.e. the state sequence is not volatile). Therefore, coefficients for exposures of interest were only interpreted for biLOGR models fitted to P&I for the 0-4, 5-14, 15-44 and 45-64 age groups and to ILI for the 0-4, 45-64 and $\geq$ 65 age groups. Volatile state sequences indicate conflict between outcome variables in estimating state sequences (section 6.7.4). The estimation of the state sequence is directly related to the estimation of the mean shift for each season (since the mean shift is the ratio of incidence in the 'aberrant' state to that in the 'normal' state). Exposures of interest act on the average mean shift (the mean of the random effect). Volatile state sequences mean unreliable estimates of the yearly mean shift, the average mean shift and therefore coefficients for exposure effects of interest.

Most triLOGR models including exposures of interest produce volatile state sequences or do not converge. Because of this, neither the effect of CT seasons nor the impact of vaccination from triLOGR model fits were interpreted.

Recall from chapter 6 that no model adequately fits P&I data for the $\geq$ 65 age group. Because of this, it was not possible to estimate effects of exposures of interest on P&I data for the $\geq$ 65 age group.

7.7. Crude results

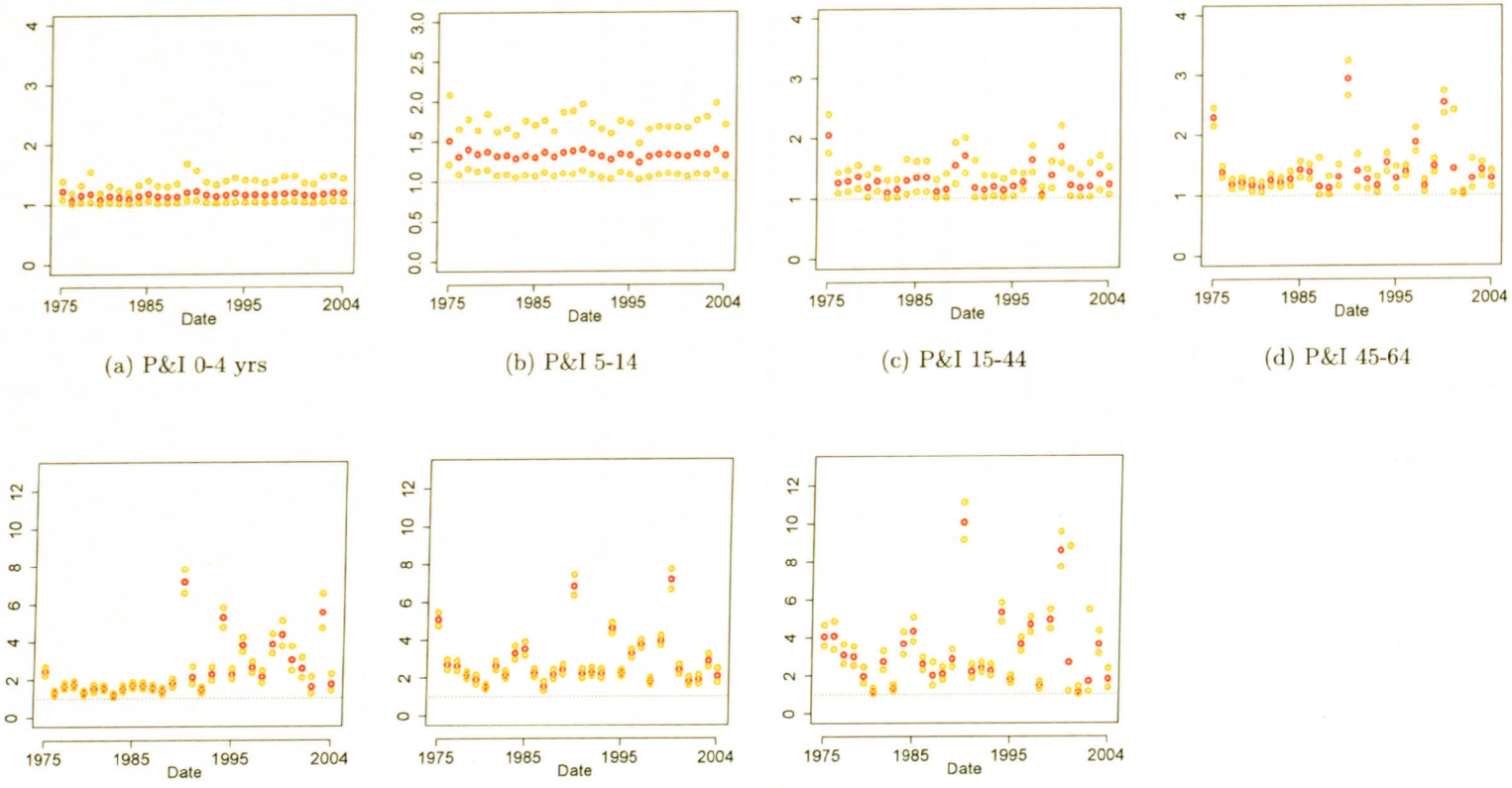
In section 7.7.1, time series of mean shifts from crude models (those not yet including exposures) were plotted to examine the variability in mean shifts between seasons, between age groups and between P&I and ILI. Recall that exponentiated mean shifts are rate ratios for the average rate in the 'aberrant' state divided by the average rate in the 'normal' state for each influenza season. In section 7.7.2, plots of the duration of the 'aberrant' period during each influenza year across models were examined for consistency between models. To create these plots, the start of the period of 'aberrant' incidence was defined as minimum two consecutive weeks designated as 'aberrant' and the end as minimum two consecutive weeks designated as 'normal'. [70]

7.7.1. Mean shifts

Plots of crude exponentiated mean shifts from models without the exposures of interest are in broad agreement across outcomes and age groups as to highest and lowest impact influenza seasons (figure 7.1). Mean shifts are more variable for ILI than for P&I.

For P&I, the distinction between high and low impact influenza seasons is clear in the 45-64 age group and less obvious in the younger age groups. This is because influenza related mortality is infrequent in younger people. 95% credible intervals (CrI) are widest for mean shifts for the 5-14 age group in which almost no deaths attributable to influenza are registered. The three highest impact (highest mean shift) seasons for the 45-64 age group are 1975/76, 1989/90 and 1999/00. 1975/76 and 1989/90 were CT seasons. Recall that the biLOGR model fitted to P&I for the $\geq$ 65 age group does not converge and so results for this age group are not shown.

For ILI, a clear distinction between high and low impact influenza seasons is visible for all age groups presented. Recall that biLOGR models fitted to ILI for the 5-14 and 15-44 age groups do not converge and so results for these age groups are not shown. 1989/90 was a high impact season in terms of ILI in all age groups presented. 1999/00 was also a high impact season for the 45-64 and $\geq$ 65 age groups. The 1975/76 season was only obviously high impact, compared with other influenza seasons, for the 45-64 age group. For the 0-4 age group, 1993/94 and 2003/04 were also notably high impact seasons.



(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64

(e) ILI 0-4 yrs (f) ILI 45-64 (g) ILI ≥ 65

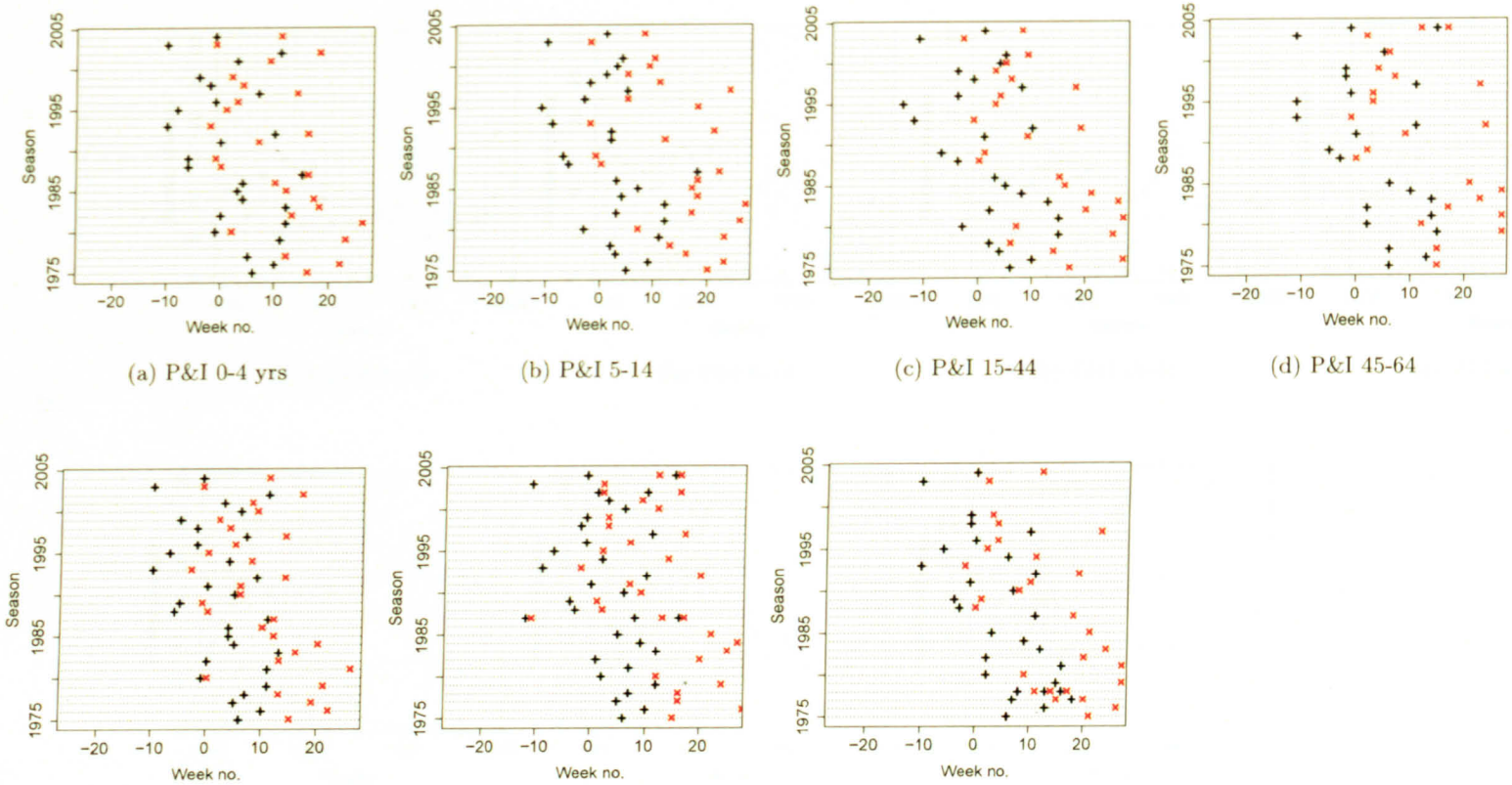
Figure 7.1.: **Time series of crude exponentiated mean shifts (red dots) and 95% CrIs (gold dots) from biLOGR models.** Horizontal dotted lines at 1 indicate the smallest possible mean shift (i.e. no difference between 'normal' and 'aberrant' incidence). Recall that mean shifts were constrained to be ≥ 0 on the logarithmic scale, ≥ 1 on the original scale. Mean shifts are only shown for models where, when exposures were added, state sequences appear to converge and are clearly estimated. There is broad agreement between models as to high mean shift (high impact) and low mean shift (mild or low impact) influenza seasons.

7.7.2. Timing and duration of 'aberrant' periods

There is broad agreement across models as to the timing of periods of 'aberrant' incidence in a given season (figure 7.2). Looking across the study period, there is a suggestion in these plots that 'aberrant' periods are starting progressively earlier (i.e. that 'aberrant' periods tend to occur later in the year in the 1970s than they do in the 1990s). This may be an artifact of the laboratory data for two reasons. First, recall from chapter 3 that the earliest specimen date is the date of report to HPA CfI, from 1975-1988, and generally the date of sample from 1989 onwards. The date of report would either be the date of sample or a later date. From 1975-1989 laboratory testing was done in batches approximately weekly. Therefore laboratory data suggesting influenza is circulating earlier in the year in influenza seasons after 1988 may simply be an artifact of the end of batch testing.

Second, up to 1993 most laboratory reports in LabBase2 were from hospitalised patients. After 1993, laboratory reports from patients visiting their GP for ILI became more prevalent in LabBase2. [33] The laboratory data from sentinel GP practices may provide an earlier indication that influenza is circulating if ascertainment of patients with influenza occurs earlier by GPs than in hospitals. It is not obvious that this should necessarily be the case because people visiting their GP with ILI and those hospitalised with influenza are probably different in several ways including health status.

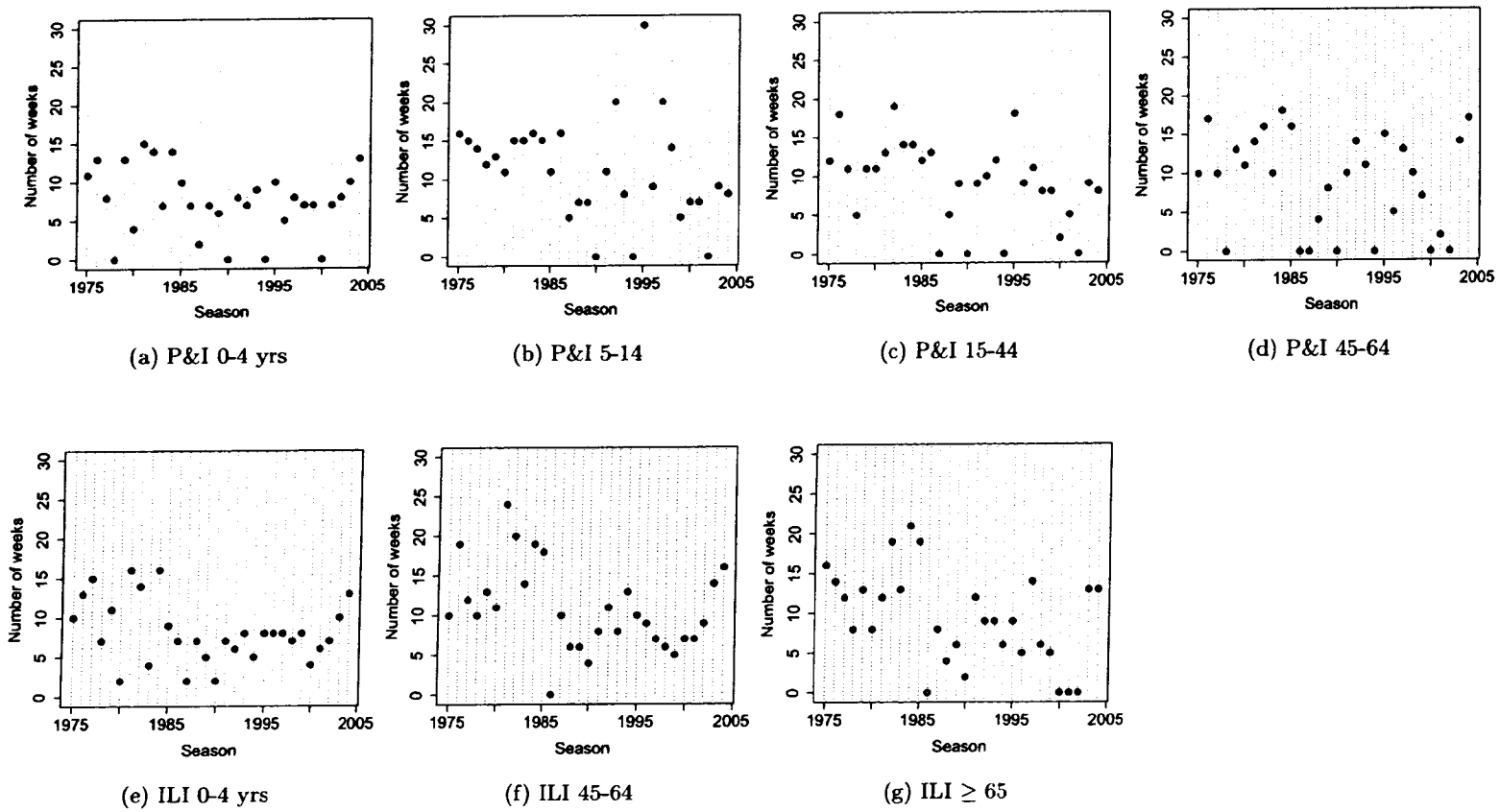
Pooling results across models without exposures of interest shows that a median 9.5 weeks (interquartile range 6 to 14 weeks) per influenza season are designated by the model as 'aberrant'. Figure 7.3 shows the high degree of variability in the duration of 'aberrant' periods influenza season to influenza season.



(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64

(e) ILI 0-4 yrs (f) ILI 45-64 (g) ILI ≥ 65

Figure 7.2.: **Estimated timing of 'aberrant' period each season from biLOGR models.** '+' denotes the start of 'aberrant' incidence; 'x' denotes the end of 'aberrant' incidence. Weeks are numbered (week 0 = week of January 1st).



(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64

(e) ILI 0-4 yrs (f) ILI 45-64 (g) ILI $\geq$ 65

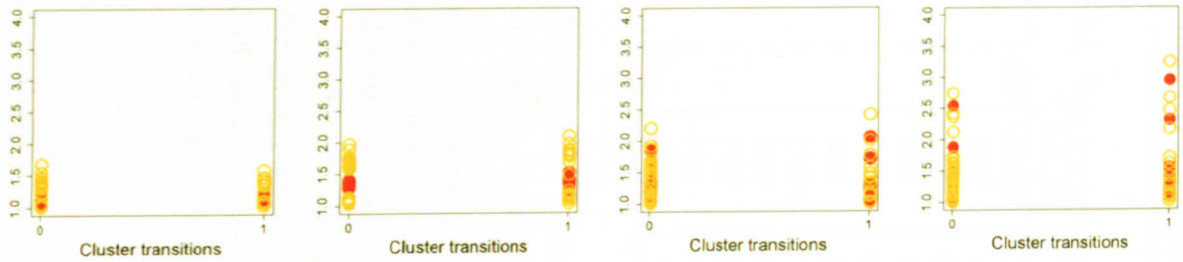
Figure 7.3.: **Estimated crude duration of 'aberrant' period each season from biLOGR models.** The points indicate the number of consecutive weeks the model estimates are in the 'aberrant' state for each influenza season. Zero denotes seasons where never ≥ 2 consecutive weeks are labelled 'aberrant'.

7.8. Effect of cluster transition seasons

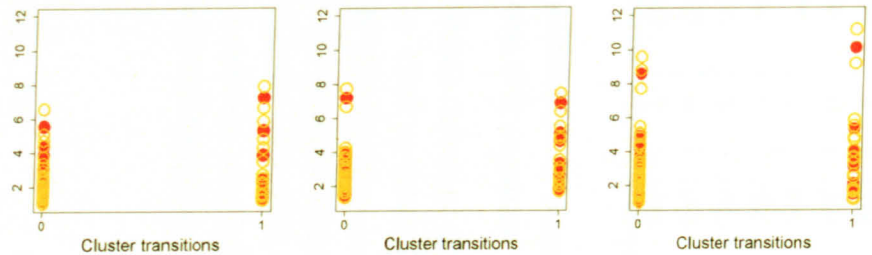
7.8.1. Association between exponentiated mean shifts and CT seasons

Mean shifts for each influenza season from models not yet including covariates (i.e. from crude models) were exponentiated and plotted against exposures of interest. Mean shifts are highly variable whether or not a CT has occurred (figure 7.4). No strong association between CTs and mean shifts is evident.

Mean shifts were ranked largest to smallest within age groups by outcome (P&I or ILI). The prominence of CT seasons in the top ten seasons was noted. Never more than 4 out of 10 of the top seasons are CT seasons (tables 7.1 and 7.2). This is consistent with findings reported in chapter 4. When peak ILI or P&I rates observed in a given season are ranked largest to smallest, CT seasons do not feature prominently in the top ten seasons.



(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64



(e) ILI 0-4 yrs (f) ILI 45-64 (g) ILI ≥ 65

Figure 7.4.: **Scatter plots of exponentiated biLOGR mean shifts against CT seasons.** Recall that CT seasons were defined as the first influenza A/H3N2 virus-dominated season after a CT. Exponentiated mean shifts are red; 95% CrIs are yellow.

Table 7.1.: **Top ten influenza seasons in terms of mean shift in P&I.** 1 is a CT season, 0 an intracluster season.

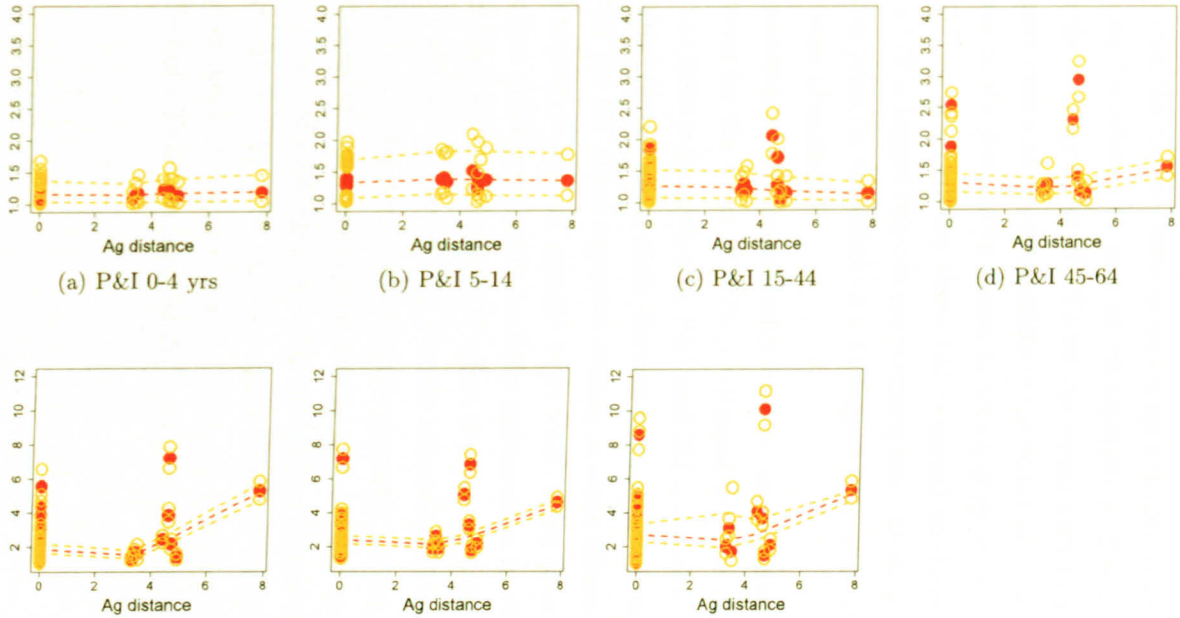
	0-4 yrs		5-14		15-44		45-64	
rank	season	CT?	season	CT?	season	CT?	season	CT?
1	1975	1	1975	1	1975	1	1989	1
2	1989	1	1977	1	1999	0	1999	0
3	1988	0	1989	1	1989	1	1975	1
4	1999	0	2003	0	1996	0	1996	0
5	2003	0	1988	0	1988	0	1993	1
6	1993	1	1979	1	2003	0	1998	0
7	1984	0	1985	0	1998	0	2000	0
8	1978	0	1987	0	1978	0	1984	0
9	2004	0	1990	0	1985	0	2003	0
10	1998	0	1978	0	1984	0	1990	0

Table 7.2.: **Top ten influenza seasons in terms of mean shift in ILI.**

	0-4 yrs		45-64		≥ 65	
rank	season	CT?	season	CT?	season	CT?
1	1989	1	1999	0	1989	1
2	2003	0	1989	1	1999	0
3	1993	1	1975	1	1993	1
4	1999	0	1993	1	1998	0
5	1998	0	1998	0	1996	0
6	1995	1	1996	0	1984	0
7	2000	0	1984	0	1976	0
8	1996	0	1995	1	1975	1
9	2001	0	1983	0	1983	0
10	1975	1	2003	0	1995	1

7.8.2. Association between exponentiated mean shifts and antigenic distance between CT seasons

It is possible that larger cluster transitions (in terms of the antigenic distance between clusters) have a greater effect on morbidity and mortality than smaller cluster transitions. Crude exponentiated mean shifts for each influenza season were plotted against the antigenic distance between clusters to determine whether, crudely, there is evidence for this in the data analyzed (figure 7.5). There is not a consistent association between exponentiated mean shifts and the size of cluster transitions.



(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64

(e) ILI 0-4 yrs (f) ILI 45-64 (g) ILI ≥ 65

Figure 7.5.: **Scatter plots of exponentiated biLOGR mean shifts vs. the antigenic distance between clusters (0 for intracluster seasons).** A lowess spline is included to visualise the association between the mean shift and antigenic distance.

7.8.3. Estimates of the effect of CTs from model fits

In this section, estimates of the effect of CT seasons on the mean shift for P&I and ILI from biLOGR model fits with dependency between the random effect mean shift and CT seasons are explored.

Models with reference priors or weakly informative priors on the effect of CTs on the mean shift are computationally difficult. For example, the two chains for the effect of cluster transitions on the mean shift with a reference or weakly informative prior do not converge for any model. The two chains for the CT coefficient appear to have difficulty visiting the parameter space (e.g. figure 7.6). Recall that in section 5.12, the relationship between information in the data and convergence of the model was introduced. The difficulty that the models have in sampling from the posterior for the effect of CTs may be related to lack of information in the data with which to estimate the average effect of CTs on the mean shift in these data.

Attempts were also made to estimate the effect of CTs in the two seasons subsequent to the first H3N2 season in which they were dominant (given no further CT occurred during these two seasons) to determine whether the effect of a CT is lagged relative to its emergence. Models with anything but a reference prior on the effect of CTs lagged by one or two seasons do not run and were abandoned.

As mentioned above, the difficulty that the models have in sampling from the posterior for the effect of CTs (illustrated in figure 7.6) indicates there may be too little information in P&I and ILI data with which to quantify the CT effect on the mean shift. Setting more informative priors on the coefficient for the CT effect improves mixing to some extent, since the models have a less diffuse area to sample from when more informative priors are set. Because there appears to be little information in the data about the CT effect, the posterior for the CT effect is essentially dictated by the prior on the CT effect (e.g. figure 7.7). For the reference prior scenario (left-most plot in figure 7.7), the two chains for the posterior for the CT effect inhabit a similar area of parameter space but do not overlap completely. This is because of lack of convergence and poor mixing of the two chains for the CT effect (figure 7.6). The variance of the posterior for the CT effect is less than the variance of the prior but is still large (encompassing a magnitude of the effect of CTs on the mean shift, on the original scale, between approx-

imately 0 and 37,000,000). The fact that the posterior is less variable (has lower variance) than the reference prior suggests there may be information in the data with which to estimate the effect of cluster transitions but that this information is limited.

For the weakly informative prior scenario (middle plot in figure 7.7), the prior and posterior are similar. The two chains for the posterior CT effect do not agree on the median CT effect (35, on the original scale, for one chain and 547 for the other), again because of lack of convergence of the two chains for the CT effect (e.g. figure 7.8). When a strongly informative prior is used, the posterior and prior are almost identical (right-most plot in figure 7.7). Both the fact that the posterior from the reference prior scenario is very vague and the sensitivity of the posterior to the choice of prior suggest that there is little information in the data which to estimate the effect of CTs on the mean shift. If there were information in the data with which to estimate the effect of CTs on the mean shift, then the variance of the posterior would be expected to be less than the variance of the reference and weakly informative priors. The variance of the posterior would be expected to encompass a plausible range of effect sizes. If there were information in the data about the effect of CTs, the estimated magnitude of this effect would be relatively robust to the choice of prior.

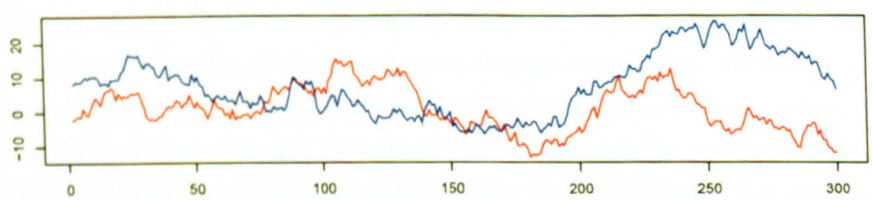
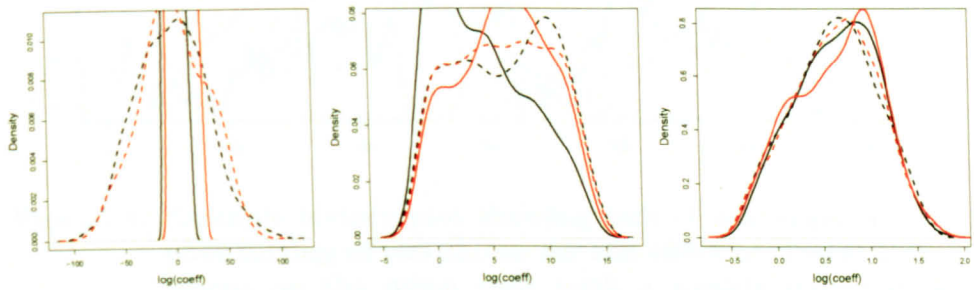


Figure 7.6.: **Example history plot showing lack of convergence and poor mixing of two chains for the effect of cluster transitions on the mean shift with a reference prior on the CT effect.** This plot is from the biLOGR fit to ILI for the $\geq$ 65 age group.

The estimate of the magnitude of the effect of each unit change in the antigenic distance between clusters on the mean shift in P&I or ILI is also sensitive to the choice of prior, though posteriors are less dominated by priors than for estimates of the effect of CTs (e.g. figure 7.9 vs. 7.7). The



(a) Reference prior (b) Weakly informative prior (c) Strongly informative prior

Figure 7.7.: **Sensitivity of the posterior for the average effect of CTs on the mean shift in ILI for the $\geq$ 65 age group to the choice of prior on the CT effect.** The y-axis is probability density and the x-axis is the CT effect on the logarithmic scale. Priors and posteriors for the two chains are plotted separately. Dotted curves are priors, solid curves are posteriors.

posterior distribution for the coefficient of the effect of each unit change in antigenic distance with a reference prior suggests a median magnitude of effect of 1.4 on the log scale (left-most plot in figure 7.9). With a weakly informative prior (middle plot in figure 7.9) two chains for the coefficient for the effect of antigenic distance do not converge, but both suggest a median effect size close to 0 on the log scale (no effect). History plots of two chains for coefficients for the effect of each unit change in antigenic distance do not mix well within the range of the posterior distribution for any prior (e.g. figure 7.10). Poor mixing of two chains for coefficients for antigenic distance reiterates that there is limited information in the P&I and ILI data with which to quantify the effect of a change in antigenic distance on mean shifts. These data cannot provide strong evidence for or against an effect. Posterior distributions for coefficients for antigenic distance are not identical to their reference or weakly informative prior distributions, meaning that the P&I and ILI data do provide some information with which to quantify the effect of different sized cluster transitions.

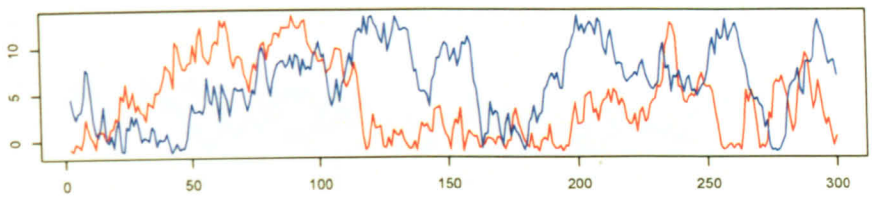
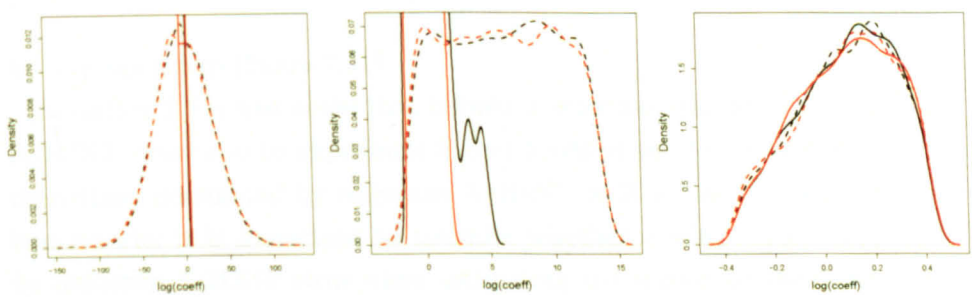


Figure 7.8.: **Example history plot showing lack of convergence and poor mixing of two chains for the effect of cluster transitions on the mean shift with a weakly informative prior.** This plot is from the biLOGR fit to ILI for the $\geq$ 65 age group.



(a) Reference prior (b) Weakly informative prior (c) Strongly informative prior

Figure 7.9.: **Sensitivity of the posterior for the average effect of each unit of antigenic distance between clusters on the mean shift in ILI for the $\geq$ 65 age group to the choice of prior on the effect of a unit change in antigenic distance.**

7.9. Impact of vaccination of the elderly on the mean shift in P&I and ILI

7.9.1. Association between exponentiated mean shifts and vaccine coverage of the $\geq$ 65 age group

Increasing vaccine coverage of the elderly each influenza season might be expected to lead to a decrease in the mean shift. Crude exponentiated mean shifts were plotted against vaccine coverage to assess whether, crudely, there is evidence for this in the P&I or ILI data. There is no clear association between increasing vaccine coverage of the elderly and the size of mean shifts

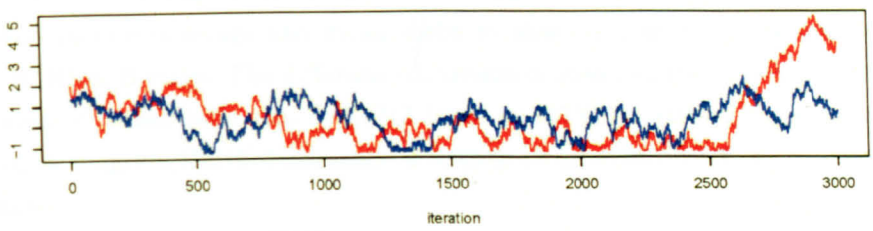


Figure 7.10.: **Example history plot showing lack of convergence and poor mixing of two chains for the effect of a unit change in the antigenic distance between clusters on the mean shift.** This plot is from the biLOGR fit to ILI for the $\geq$ 65 age group with a weak prior on the effect of a unit change in the antigenic distance between clusters.

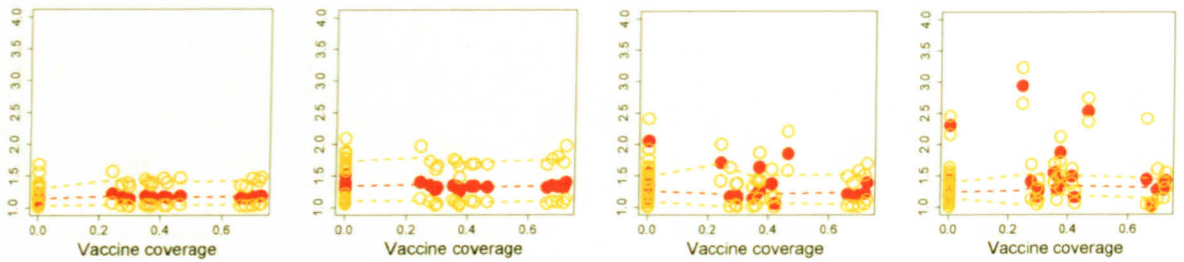
for any age group (figure 7.11).

In section 2.5 it was noted that influenza seasons dominated by influenza A/H3N2 virus tend to experience higher levels of morbidity and mortality than those dominated by influenza A/H1N1 or B viruses. There are two reasons why it is important to consider whether seasons were dominated by influenza A/H3N2 virus when estimating the impact of vaccination on trends in the relative impact of influenza seasons (here, the mean shift). Influenza A/H3N2 virus-dominance may act as, first, an effect modifier of vaccine impact and/or, second, a confounder of vaccine impact.

First, influenza A/H3N2 virus-dominance may act as an effect modifier of vaccine impact. VE may be less during lower impact influenza seasons. This is because it is expected that morbidity or mortality occurring during lower impact seasons is less specific to influenza. Morbidity and mortality not attributable, or secondary, to influenza would be expected to be distributed evenly between vaccinated and unvaccinated people (assuming no difference in the underlying health status of vaccinated and unvaccinated people). Lower VE during lower impact influenza seasons would be expected to result in a smaller vaccine impact in lower impact influenza seasons. In order to determine whether, crudely, the association between mean shifts and vaccination coverage is different depending on whether influenza A/H3N2 viruses were dominant, mean shifts were colour-coded as H3N2-dominated seasons (red/gold) or not (blue/green) on the scatter plot of mean shifts against vaccine coverage (figure 7.12). There is not a consistent associa-

tion between vaccine coverage and mean shifts in H3N2-dominated seasons. For H3N2-dominated seasons mean shifts appear to be higher during influenza seasons with moderate vaccine coverage than in seasons with either no vaccine coverage or high vaccine coverage. There is no association between vaccine coverage and mean shifts in seasons dominated by influenza A/H1N1 or B virus. The difference in trends in mean shifts for seasons dominated by either influenza A/H3N2 virus or H1N1/B viruses for some age groups suggests H3N2-dominance is acting as an effect modifier of vaccine impact.

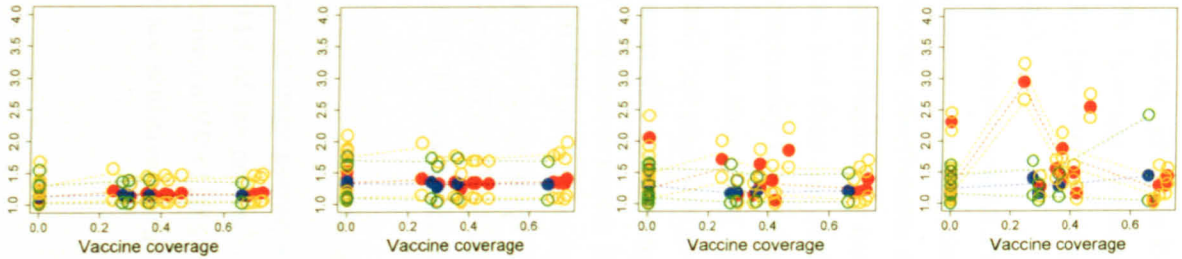
Second, influenza A/H3N2 virus-dominance may act as a confounder of vaccine impact. If, for example, an increase in the frequency of influenza A/H3N2 virus-dominated seasons had occurred towards the end of the study period, when vaccine coverage was highest, this might have caused negative confounding of the impact of vaccine on the mean shift. Influenza A/H3N2 virus-dominated seasons are not more or less frequent during the period when vaccine coverage was moderate or when coverage was high. It is unlikely that influenza A/H3N2 virus-dominance is acting as a confounder of vaccine impact.



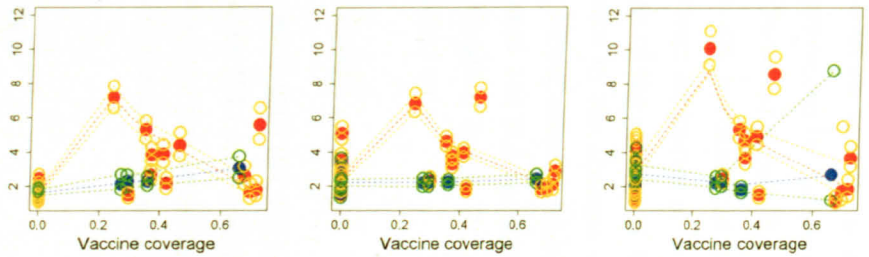
(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64

(e) ILI 0-4 yrs (f) ILI 45-64 (g) ILI ≥ 65

Figure 7.11.: **Scatter plot of exponentiated biLOGR mean shifts (red dots) and their 95% CrIs (golden dots) for P&I (top) and ILI (bottom) vs. vaccine coverage in those ≥ 65.** A lowess spline is included to visualise the association between the mean shift and vaccine coverage.



(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64



(e) ILI 0-4 yrs (f) ILI 45-64 (g) ILI ≥ 65

Figure 7.12.: **Scatter plot of exponentiated biLOGR mean shifts and their 95% CrIs vs. vaccine coverage in those ≥ 65, stratified by whether the influenza season was an influenza A/H3N2 virus-dominated season (red/gold) or not (blue/green).** A lowess spline is included to visualise the association between the mean shift and vaccine coverage.

7.9.2. Expected vaccine impact

In this section, the expected impact of vaccination of the $\geq$ 65 age group during moderate and high coverage periods on excess ILI and on excess P&I, both in the $\geq$ 65 age group, is posited. The expected impact of vaccination on each outcome is based on estimates of average effectiveness of trivalent inactivated influenza vaccination of the elderly from the literature, vaccine coverage and presumed average cumulative excess ILI and P&I in influenza seasons before appreciable vaccine coverage was achieved. Presumed average cumulative excess ILI/P&I before vaccination coverage was appreciable captures what excess would be, on average, if vaccine coverage were zero. Two levels of average vaccine coverage were used: 36% (the average coverage of the $\geq$ 65 age group in England & Wales between 1989/90 and 1999/00) and 70% (average coverage of the $\geq$ 65 age group between 2000/01 and 2004/05) (see the following section for derivations of these cut-points in coverage).

Recall from section 2.8.2 that best estimates of vaccine effectiveness against acute respiratory hospitalisations in the elderly are between 20 and 30%. [114, 216] Vaccine effectiveness against ILI would be expected to be lower than that against respiratory hospitalisations because ILI may be less specific to influenza and thus less likely to be prevented by influenza vaccination than acute respiratory hospitalisations. Presumed excess ILI rates in England & Wales in the absence of vaccination are hypothetical (since no paper was identified) but plausible given the range estimates of excess ILI rates during seasons with low vaccine coverage of the elderly. [1, 65] Table 7.3 shows that if cumulative excess ILI incidence in the absence of vaccination were 2000/100,000 per influenza season, during seasons when VE against ILI is 20% and coverage (p) is 36%, at most 144 ILI consultations per 100,000 people over 65 (7.2% of the pre-vaccination excess ILI rate) (d_a) could be prevented. This estimate was calculated as

$$d_a = (2000/100,000) * VE * p$$

When coverage averages 70%, at most 280/100,000 excess ILI consultations (14% of the pre-vaccination excess ILI rate) would be prevented each season given a VE of 20%. Estimates of the predicted impact of vaccination on ILI are sensitive to the presumed average excess ILI rate before vacci-

nation. These estimates are the impact due to direct effects of vaccination only. Assumptions underlying these calculations are that vaccinated and unvaccinated elderly are similar in all ways related to risk of morbidity or mortality apart from vaccination status and that VE is constant influenza season to influenza season. The calculation is sensitive to the presumed excess morbidity and mortality in the absence of influenza.

Even though it was not possible to fit HMMs to P&I for the $\geq$ 65, it is worth noting the range of possible rates of excess P&I prevented given the vaccine coverage achieved and estimates of VE against P&I from the literature. In section 2.8.2 best estimates of VE against respiratory mortality in the elderly between 12% (95% CI 8-16%) and 79% (0-100%) were noted. [77, 113] These estimates were taken to be the best estimates of VE against P&I. The average excess P&I rates in influenza seasons before much influenza vaccine was distributed to the elderly are based on estimates from the US [7] and France [79]. When vaccine coverage was 70%, assuming VE of 79% against P&I and average pre-vaccination incidence of excess P&I of 262/1,000,000 per influenza season, an average of up to 145 deaths per 1,000,000 people aged $\geq$ 65 each influenza season (55% of pre-vaccination excess P&I) might have been prevented (table 7.4).

The estimates in tables 7.3 and 7.4 are of the impact in the elderly attributable only to the direct effect of some of them having been vaccinated. These estimates of vaccine impact do not include the possible indirect effects afforded to unvaccinated elderly attributed to their mixing among vaccinated elderly. It is unlikely that the elderly play a strong role in the community transmission of influenza since transmission appears greatest from children and teenagers. [38, 39] If there were an impact in other age groups due to the vaccination of the elderly - an indirect effect of vaccinating the elderly in dampening transmission of influenza in the community among other age groups - it would be expected to be much smaller than the (direct) impact in the elderly in tables 7.3 and 7.4. A small impact in other age groups would be difficult to detect and to quantify using modeling. As a first step, an attempt was made to estimate the magnitude of vaccine impact in the elderly.

Table 7.3.: **Range of possible impact of vaccination of the elderly on excess ILI in the elderly.**

Presumed average cumulative excess ILI per 100,000 in the ≥ 65 age group each season in the absence of vaccination	**2000**	**1500**	**1000**	**500**
Estimated excess ILI per 100,000 in ≥ 65 prevented each season assuming 36% coverage and 20% protection (36% coverage and 30% protection)	144 = 2000 x 0.2 x 0.36 (216)	108 (162)	72 (108)	36 (54)
Estimated excess ILI per 100,000 in ≥ 65 prevented each season assuming 70% coverage and 20% protection (70% coverage and 30% protection)	280 (420)	210 (315)	140 (210)	70 (105)

Table 7.4.: **Range of possible impact of vaccination of the elderly on excess P&I in the elderly.**

Presumed average cumulative excess P&I per 1,000,000 in the ≥ 65 each season in the absence of vaccination	**262** [6]	**148**[7]
Estimated excess P&I per 1,000,000 in ≥ 65 age group prevented each season assuming 36% coverage and 12% protection (36% coverage and 79% protection)	11 (75)	6 (42)
Estimated excess P&I per 1,000,000 in ≥ 65 age group prevented each season assuming 70% coverage and 12% protection (70% coverage and 79% protection)	22 (145)	12 (82)

[6] average excess P&I rate between 1969/70 and 1972/73 in the US, fig 3B from [7]
[7] average excess P&I between 1980/81 and 1984/85 in France, table 2 from [79]

7.9.3. Estimates of vaccine impact on mean shifts in ILI in the ≥ 65 age group from model fits

In this section, results of biLOGR model fits to estimate the impact of vaccination on the mean shift in ILI in the ≥ 65 age group are summarised. Models of the impact of each unit increase in vaccine coverage (i.e. vaccine coverage as a quantitative variable) on the mean shift with either reference or weakly informative priors on vaccine impact are computationally problematic. The two chains for the coefficient for vaccine impact with a reference or weakly informative prior do not converge for any model. The two chains for the coefficient for vaccine impact appear to have difficulty visiting the parameter space (e.g. figure 7.13).

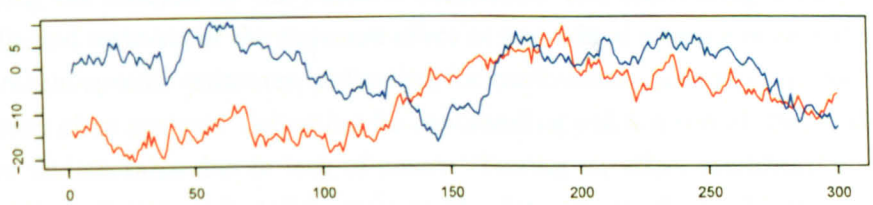


Figure 7.13.: **Example history plot showing lack of convergence and poor mixing of two chains for the impact of vaccination on the mean shift.** This plot is from the biLOGR fit to ILI for the ≥ 65 age group with a reference prior on the impact per unit increase in vaccine coverage.

BiLOGR models were also fitted to ILI for the ≥ 65 age group with a dependency between the mean shift and vaccine coverage as a categorical variable. This model compared the average mean shift in seasons between 1975/76 and 1988/89 (the reference period when vaccine coverage of the ≥ 65 age group was assumed to be 0) to the average mean shift in seasons with moderate vaccine coverage of the elderly (average 36% coverage between 1989/90 and 1999/00) and to the average mean shift in seasons with high vaccine coverage of the elderly (average 70% coverage between 2000/01 and 2004/05).

Posterior distributions for the impact of vaccination in the ≥ 65 age group are dominated by their respective priors (e.g. figure 7.14). This suggests that there is little information in the ILI data for the ≥ 65 age group with which to estimate the impact of moderate or high vaccine coverage on the

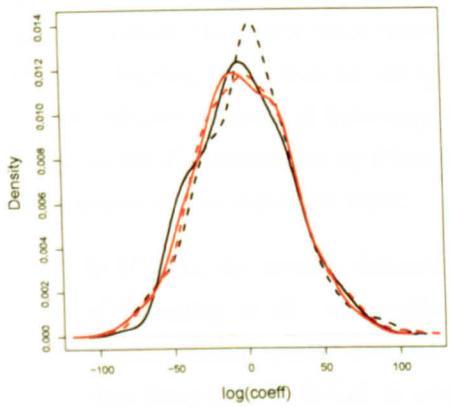
mean shift.

7.10. Confounding and effect modification

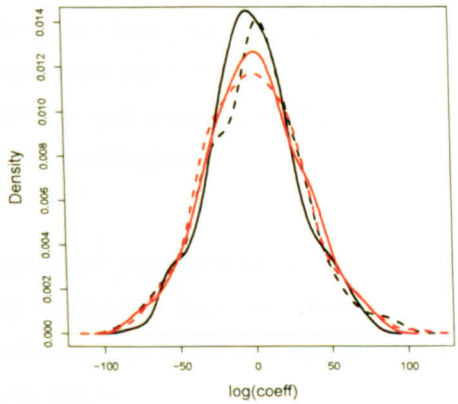
Since there appears to be little information in the data with which to estimate exposure effects of interest, there is no impetus to include confounders in these models or to stratify by possible effect modifiers (such as H3N2-dominance, in the case of estimating vaccine impact). For attempts to estimate the effect of CT seasons or vaccine impact, the variance of the posterior distributions is diffuse and encompasses a range of implausible estimates of the exposure effects of interest. For attempts to estimate the effect of a unit change in the antigenic distance between clusters on the mean shift, thereby allowing for the different size of CTs in terms of antigenic distance, two chains for coefficients for the effect of antigenic distance do not mix well, indicating low power. Adjusting for confounding involves stratifying the analysis by the putative confounder and calculating a summary adjusted estimate of the exposure effect as a weighted average of confounder stratum-specific estimates. Adjusting for confounders does not increase the power of an analysis. Adjusting for confounding will not reveal effects which are undetectable due to lack of power. Testing for effect modification also requires that there be information in the data about effects of interest.

7.11. Summary of results

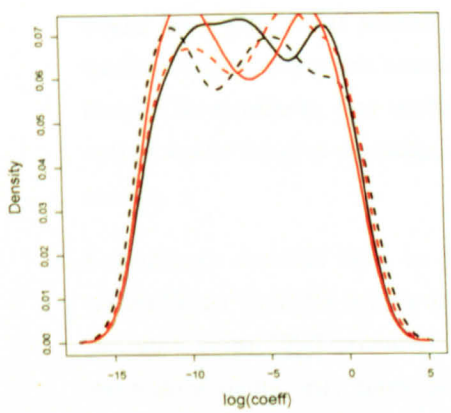
Scatter plots of crude mean shifts against covariates suggest little evidence of association between exposures of interest and mean shifts in P&I or ILI. There is limited information in the data about vaccine impact or an effect of CTs on mean shifts, so posteriors for vaccination and for CT seasons are little influenced by the data and more influenced by priors. Analysing the effect of each unit change in antigenic distance between clusters is more powerful than analysing CT seasons as a binary variable (the latter does not allow for some CTs being bigger than others in terms of antigenic units). Even allowing for the different size of CTs, there is limited information in the data about the effect of CTs on the size of epidemics.



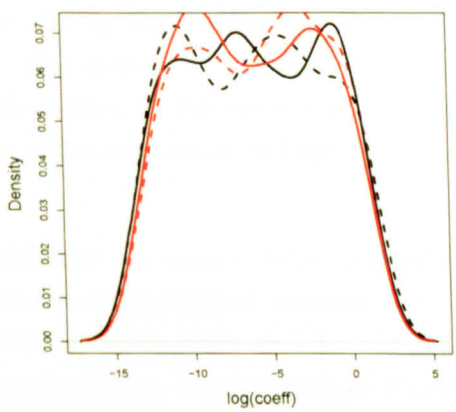
(a) Reference prior - moderate coverage



(b) Reference prior - high coverage



(c) Weakly informative prior - moderate coverage



(d) Weakly informative prior - high coverage

Figure 7.14.: **Sensitivity of the posterior for the impact of moderate or high vaccine coverage on the mean shift in ILI for the $\geq$ 65 age group to the choice of prior on the coefficients for vaccine impact.**

7.12. Results in context

In terms of ILI, the relative impact of influenza seasons in England & Wales using mean shifts estimated from biLOGR HMMs is comparable to estimates of the relative impact of influenza seasons using excess ILI rates (figures 7.15, 7.16 and 7.17). [1–3] Recall from section 2.3.1 that early estimates by Fleming *et al.* (black circles in figures 7.15, 7.16 and 7.17) [1], the black and gray "+" in figure 7.15 [2]) were derived differently from later estimates (black diamonds in figures 7.15 and 7.17 [3]). Therefore comparisons between the work undertaken for the thesis and estimates of excess ILI by Fleming *et al.* will be made separately for the two methods.

The relative impact of influenza seasons was comparable for HMMs and early excess ILI estimates by Fleming and colleagues [1,2] despite a few key differences in the methods used:

1. In HMMs, the models defined 'aberrant' periods whereas in the method of Fleming *et al.* an 'epidemic threshold' was calculated, and 'influenza active weeks' defined, externally to model fitting (this method was described in detail in section 2.3.1).

2. In the HMM analysis, the estimate of the relative impact of influenza seasons was the mean shift: the ratio of the average rate in 'aberrant' weeks for a particular season divided by the average rate in 'normal' weeks for that influenza season. The method of Fleming *et al.* calculated RRs similarly, but multiplied them by the number of 'influenza active weeks' to give an estimate of the total excess ILI rate for a given season.

3. Laboratory reports held in LabBase2 are mostly from hospitalised patients up to 1993 and a mixture of hospitalised patients and GP patients from 1993 onwards. The analyses performed for the thesis were done using only these laboratory data. The analyses by Fleming and colleagues were based on laboratory data which include the data in LabBase2 and additional laboratory data from joint HPA CfI - RCGP sentinel GP surveillance [34], results of which are not stored in LabBase2.

4. Fleming *et al.* used laboratory data aggregated by age and influenza

type/subtype. If there are different 'aberrant'/'influenza active' periods for different age groups, the method would not account for this. In the analysis described in the thesis, age group-specific influenza A laboratory reports were used, thus allowing for different timing of 'aberrant' periods between age groups.

5. Excess ILI from Fleming's 2005 paper (black and gray "+" in figure 7.15) was attributed to either influenza or RSV (see detail in section 2.3.1). BiLOGR mean shifts estimated jointly from ILI/P&I and laboratory data were attributed solely to influenza.

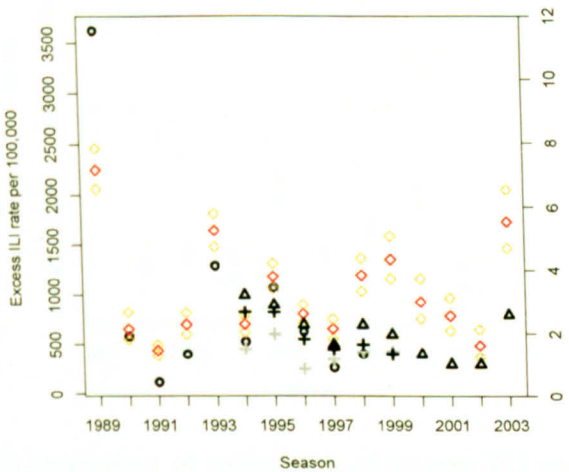


Figure 7.15.: **Comparison of estimates of excess ILI rates in children <5 in England & Wales (right axis) with exponentiated mean shifts estimated by HMMs (left axis).** Black/gray symbols are excess ILI estimates as follows: black circles are excess ILI in the 0-4 age group from [1], gray "+" are excess ILI in <1 age group and black "+" are excess ILI in the 1-4 age group from [2], black diamonds are excess ILI in the 1-4 age group from [3]. Red symbols are mean shifts for ILI in the 0-4 age group, with 95% CrIs (gold).

For the later Fleming paper, excess ILI was attributed to either influenza virus or RSV using a different method of defining 'influenza (or RSV) active periods' (figures 7.15 and 7.17). [3] The definition of 'active periods' for either virus was the average ILI each winter in the weeks (surrounding the

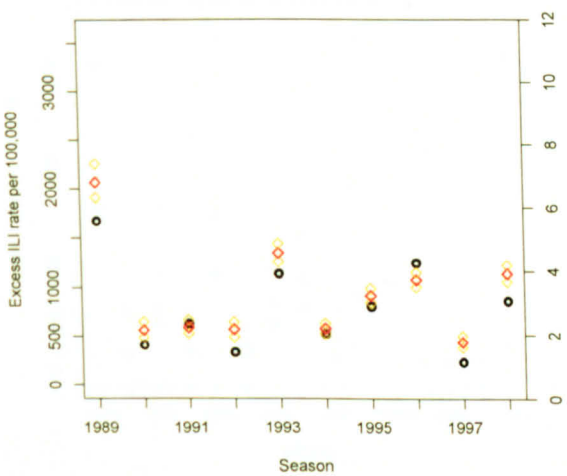


Figure 7.16.: **Comparison of estimates of excess ILI rates in the 45-64 age group in England & Wales (right axis) with exponentiated mean shifts estimated by HMMs (left axis).** Black circles are excess ILI in the 45-64 age group from [1]. Red symbols are mean shifts for ILI in the 45-64 age group, with 95% CrIs (gold).

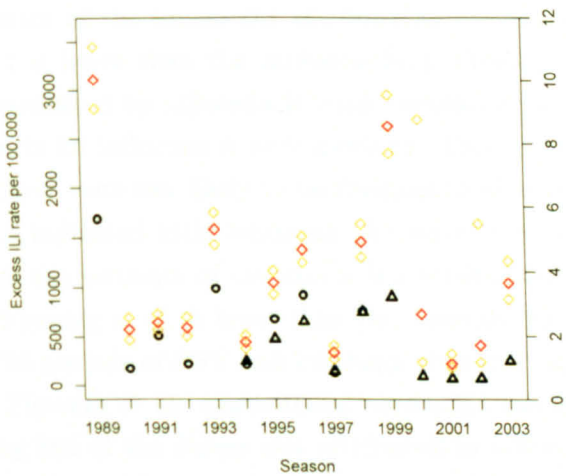


Figure 7.17.: **Comparison of estimates of excess ILI rates in the elderly in England & Wales (right axis) with exponentiated mean shifts estimated by HMMs (left axis).** Black circles are excess ILI in the $\geq$ 65 age group from [1] and black diamonds are excess ILI in the 65-74 age group from [3]. Red symbols are mean shifts for ILI in the $\geq$ 65 age group, with 95% CrIs (gold).

peak week of age-aggregated routine laboratory reports) that encompassed 70% of all laboratory reports for either influenza or RSV that influenza season. The baseline for calculation of excess was the winter period encompassing fewer than 30% of the laboratory reports for either virus. For periods where influenza and RSV circulation overlapped, excess ILI during those weeks of that winter was apportioned to RSV and influenza proportional to the excess for that winter that was attributed to either virus alone. For three influenza seasons for which there are estimates from the HMMs and from Fleming and colleagues using this method, the two methods disagree on the relative impact of the 1994/95 and 1995/96 influenza seasons for children <5 (figure 7.15). Fleming estimated that 1995/96 was a slightly lower impact season than 1994/95 whereas the HMM estimated that 1995/96 was clearly higher impact than 1994/95. There are two reasons for this discrepancy. First, the estimate of the excess ILI attributable to influenza in 1994/95 from the HMM is lower than the estimate from Fleming *et al.* because 1994/95 was dominated by influenza B virus circulation; in the HMM only laboratory reports for influenza A were modeled. This means that weeks in the 1994/95 season were less likely to be designated as 'aberrant' since the laboratory data indicated little influenza (A) activity during the 1994/95 season. Second, the estimate of the excess ILI attributable to influenza in 1995/96 from Fleming *et al.* is lower than the estimate from the HMM because in 1995/96 periods of RSV and influenza virus circulation overlapped considerably. Fleming *et al.* apportioned excess ILI incidence to the two viruses meaning less of the excess was attributed to influenza. This had a proportionately bigger effect on the estimate of excess ILI in children <5 than on estimates for adults $\geq$ 65 (figure 7.17).

Estimates of relative impact of influenza seasons in England & Wales in terms of excess respiratory mortality rates from the literature are comparable to estimates of the relative impact of influenza seasons in terms of mean shifts in P&I from HMMs (figures 7.18). This is despite differences in the method of defining relative impact of influenza seasons in terms of mortality: in HMM fits, 'aberrant' weeks were defined separately for ILI and for P&I; Fleming defined an epidemic threshold, and 'influenza active weeks', using ILI and laboratory data and used the same 'influenza active weeks' to define excess mortality.

For children, variability influenza season to influenza season in excess

respiratory mortality in the age group 1-4 yrs is low and is similar to the variability influenza season to influenza season in mean shifts in P&I for children 0-4 (figures 7.18 to 7.20). [57] Fleming and colleagues found the variability in excess respiratory mortality influenza season to influenza season is greater for the age group <1; [57] this age group was not modeled for the thesis. Weekly rates of laboratory reports for children <1 are extremely low, limiting the ability to use multivariate HMMs to estimate mean shifts for this age group.

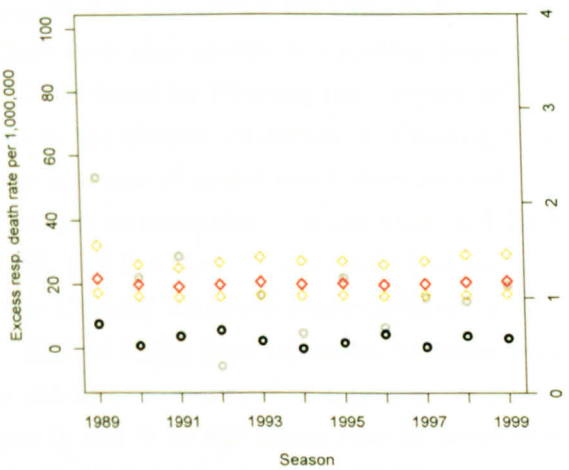


Figure 7.18.: **Comparison of estimates of excess respiratory mortality rates in children <5 in England & Wales (right axis) with exponentiated mean shifts estimated by HMMs (left axis).** Black circles are excess respiratory mortality (pneumonia, influenza or bronchitis) in the 0-4 age group, and gray circles are excess respiratory mortality in <1 age group, both from [57]. Red symbols are mean shifts for P&I in the 0-4 age group, with 95% CrIs (gold).

For people 45-64, the relative impact of influenza seasons in terms of excess respiratory mortality in England & Wales from the literature and mean shifts in P&I are similar (figure 7.19). [53] The 1996/97 influenza season was higher impact than the 1997/98 season; this distinction was more pronounced for the method of Fleming and colleagues than for the HMM method. In the 1996/97 season, influenza A circulated in the first

part of the season and influenza B circulated in the latter part. [6] This kind of longer duration, mixed type, influenza season could give a higher estimate of excess mortality compared to the mean shift in P&I estimated using the HMM. There are two reasons for this. First, the mean shift does not take into account the duration of the influenza season: the mean shift is simply the ratio of the average rates in 'aberrant' to 'normal' weeks for a given influenza season. The estimate of excess rates of mortality by Fleming *et al.* were rate ratios (similar to mean shifts), multiplied by the number of 'virus active weeks' in the season. A long influenza season can result in high estimates of excess mortality (because the excess is accrued in each week for many weeks) without affecting the mean shift. Second, in the HMM, the mean shift was only estimated using influenza A laboratory reports, so mortality occurring after the end of circulation of influenza A would is downweighted in estimating the state sequence and thus the mean shift. Both of these facts may explain the greater disparity between seasons 1996/97 and 1997/98 found by Fleming than estimated using the HMM.

For the deaths in the elderly, estimates by Fleming *et al.* of the relative impact of seasons in terms of excess respiratory mortality for the 65-74 age group are comparable to estimates of mean shift in P&I for the $\geq$ 65 age group (figure 7.20). [53] For the $\geq$ 75 age group (not modeled for the thesis), the impact of the 1996/97 influenza season relative to the 1997/98 season is pronounced. Risk of dying from influenza increases with advancing age, because of the prevalence of comorbid illnesses, so the relative impact of influenza seasons in the $\geq$ 75 age group may be more specific to influenza than the relative impact in the $\geq$ 65 or 65-74 age groups (whose deaths might be more evenly distributed throughout the year, though still higher in winter than summer).

The timing of 'influenza active periods' as defined by the two methods of Fleming and colleagues discussed above - based on an 'epidemic threshold' defined using ILI in weeks when there were no laboratory reports for influenza [1,2] (figure 7.21) and using weeks surrounding the peak laboratory report week that encompassed at least 70% of the laboratory reports for that season [3] (figure 7.22) - were similar to the timing of 'aberrant' periods estimated using the HMM.

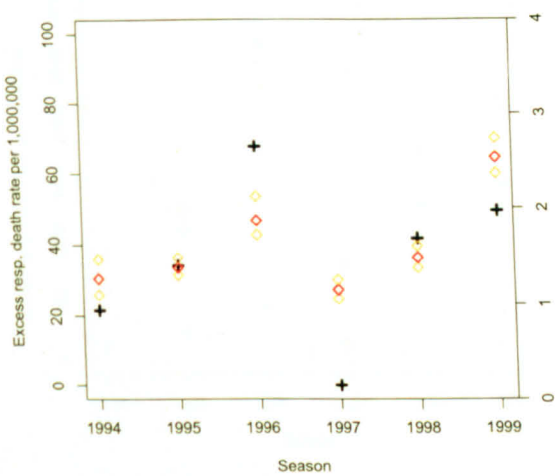


Figure 7.19.: **Comparison of estimates of excess respiratory mortality rates in adults 45-64 in England & Wales (right axis) with exponentiated mean shifts estimated by HMMs (left axis).** Black '+' are excess respiratory mortality (pneumonia, influenza or bronchitis) in the 45-64 age group from [53]. Red symbols are mean shifts for P&I in the 45-64 age group, with 95% CrIs (gold).

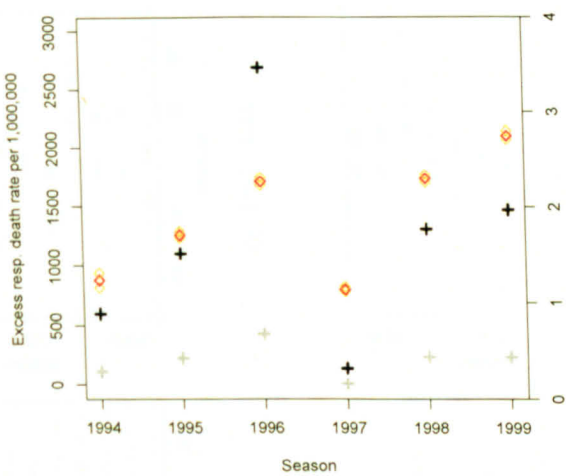
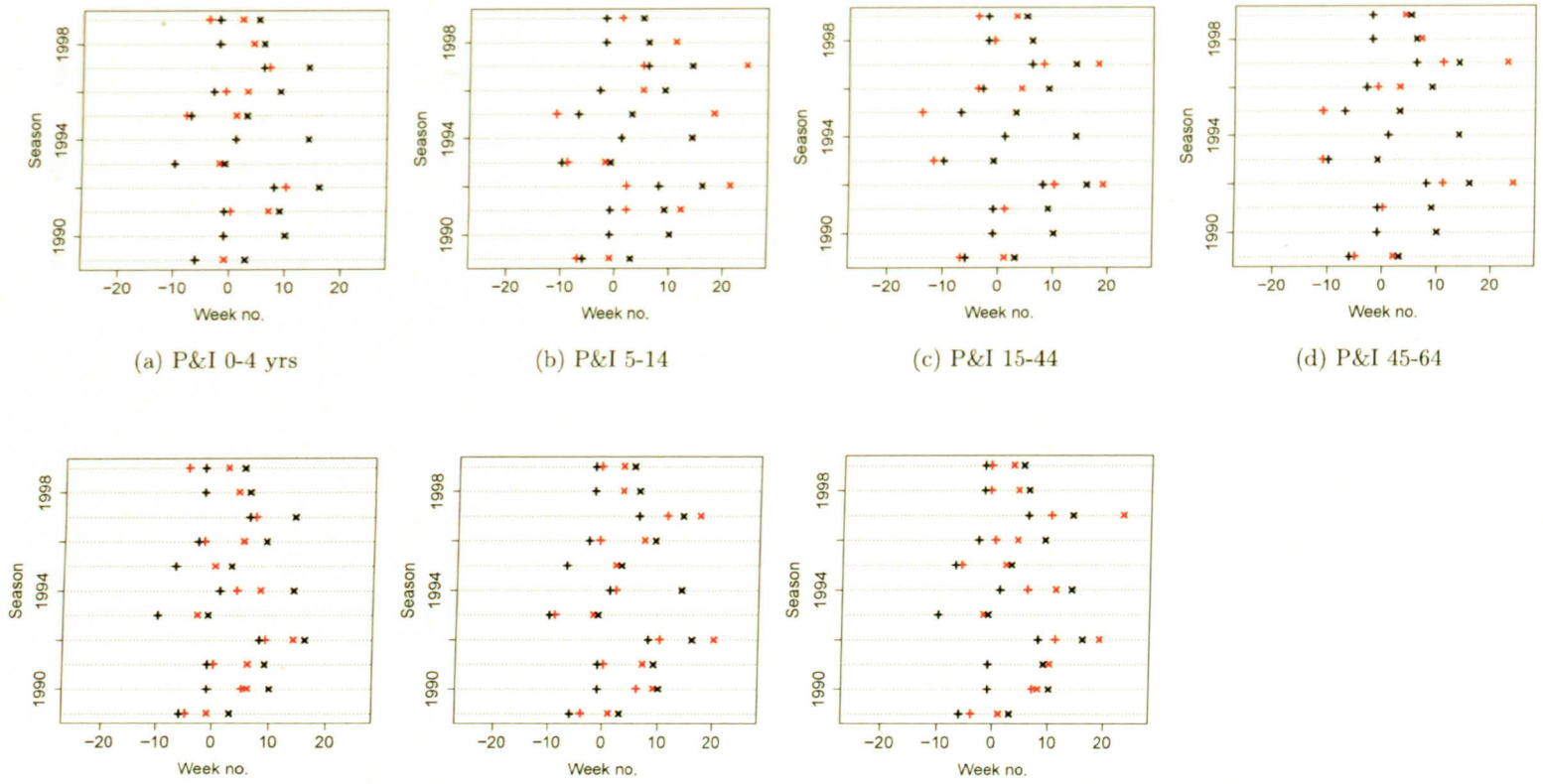


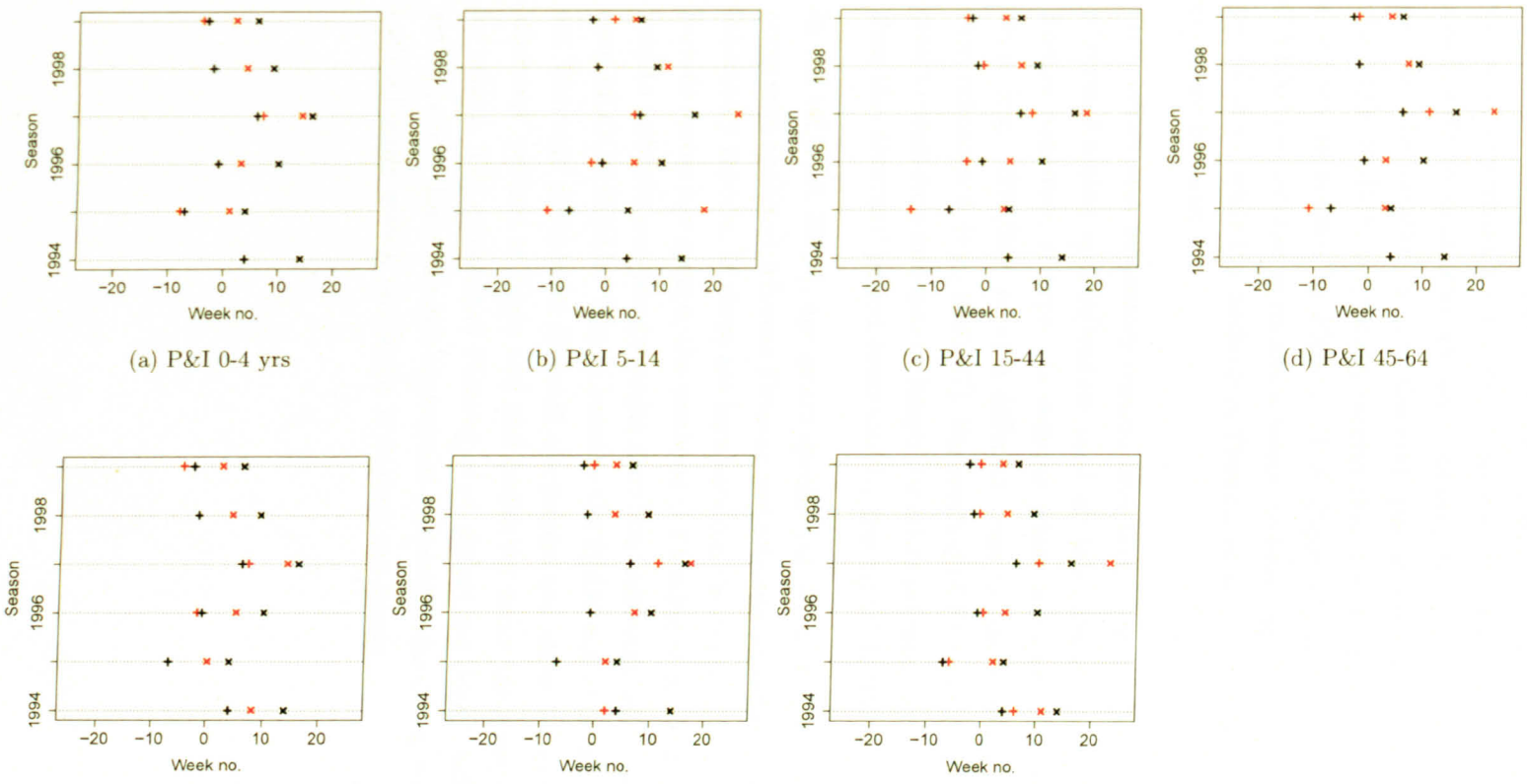
Figure 7.20.: **Comparison of estimates of excess respiratory mortality rates in the elderly in England & Wales (right axis) with exponentiated mean shifts estimated by HMMs (left axis).** Black '+' are excess respiratory mortality (pneumonia, influenza or bronchitis) in the $\geq$ 75 age group, and gray '+' are excess respiratory mortality in the 65-74 age group, both from [53]. Red symbols are mean shifts for P&I in the $\geq$ 65 age group, with 95% CrIs (gold).



(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64

(e) ILI 0-4 yrs (f) ILI 45-64 (g) ILI ≥ 65

Figure 7.21.: **Comparison of timing of 'aberrant' periods for England & Wales estimated using the method of [1, 2] (black symbols) and using the HMM (red symbols).** In [1, 2], 'aberrant' (or 'influenza active') periods were defined as weeks when observed all-age ILI was greater than the epidemic threshold for that week. The same 'influenza active' periods were applied to each age group and outcome to estimate excess. The epidemic threshold was defined as the upper 95% CI on the average ILI in weeks when there were no laboratory reports for influenza. Weeks are numbered (week 0 = week of January 1st).



(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64

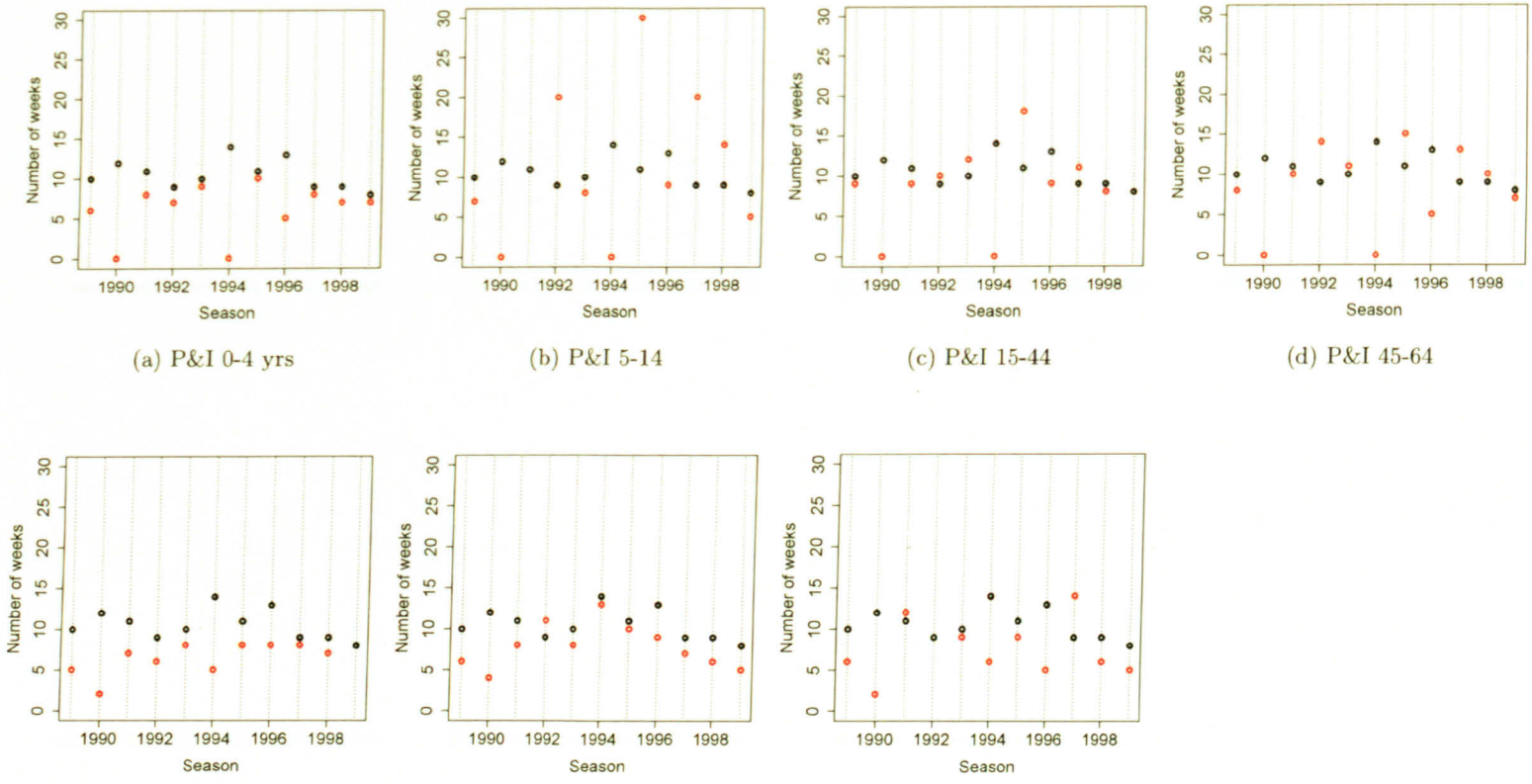
(e) ILI 0-4 yrs (f) ILI 45-64 (g) ILI $\geq$ 65

Figure 7.22.: **Comparison of timing of 'aberrant' periods for England & Wales estimated using the method of [3] (black) and using the HMM (red).** In [3], 'aberrant' (or 'influenza active') periods were defined as weeks surrounding the peak week of influenza laboratory reports that encompass at least 70% of all influenza laboratory reports (all-age) for that influenza season.

The median duration of the 'aberrant' period each influenza season estimated using biLOGR HMMs (9.5 weeks, interquartile range 6-14 weeks) is comparable to the mean length of the influenza season, 10 weeks, estimated from French ILI data. [86] Rath *et al.* estimated the mean duration of the influenza season by fitting 2 state HMMs to weekly sentinel surveillance for influenza-like-illness data from France between the 1975/76 and 1996/97 seasons. [86] The duration of the 'aberrant' period each season estimated from the biLOGR HMM is typically shorter than the duration of 'influenza active periods' as defined by Fleming [1–3] (figures 7.23 and 7.24). The biLOGR HMM sometimes estimates a season to have experienced no 'aberrant' period at all, while the methods of Fleming *et al.* never does so. These differences are because:

1. Only influenza A laboratory reports were used to fit biLOGR HMMs, whereas Fleming and colleagues used all laboratory reports for influenza regardless of type. In seasons where influenza B circulated alone (e.g. 1994/95), or at a different time in the influenza season than influenza A (e.g. 1996/97), the method of Fleming *et al.* would tend to estimate a non-zero or longer duration 'influenza active' period than the 'aberrant' period estimated by the biLOGR HMM.

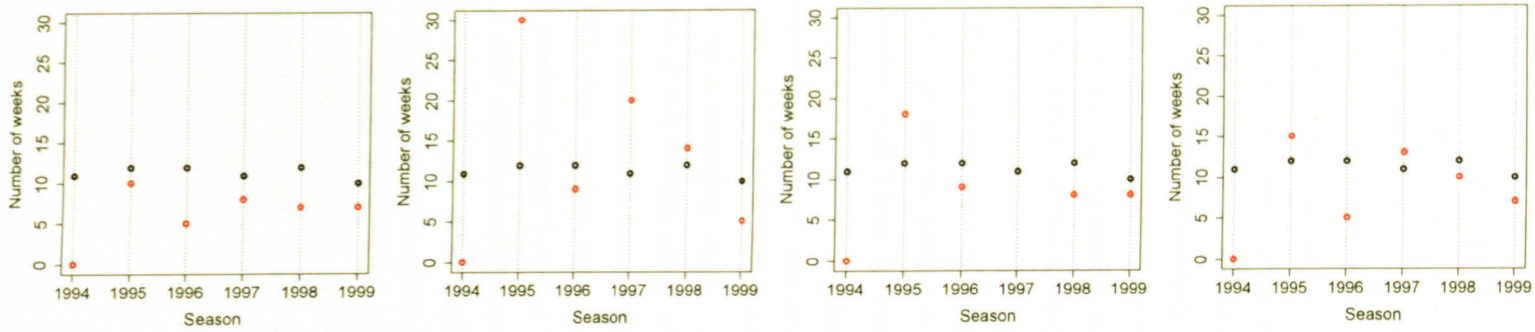
2. In the biLOGR HMM, age group-specific influenza A laboratory reports were analyzed, whereas Fleming and colleagues used age-aggregated laboratory reports. If there are lags between the timing of 'aberrant' periods across age groups, the methods of Fleming *et al.* would again tend to estimate non-zero or longer duration 'influenza active' periods than biLOGR HMMs. This is because the HMM would define different 'aberrant' periods for different age groups and allow, for example, 'aberrant' periods to begin and end earlier for some age groups than for others. The methods of Fleming *et al.* define one (long) 'influenza active period' from the age-aggregated laboratory (and ILI) data for each influenza season and apply it to each age group.



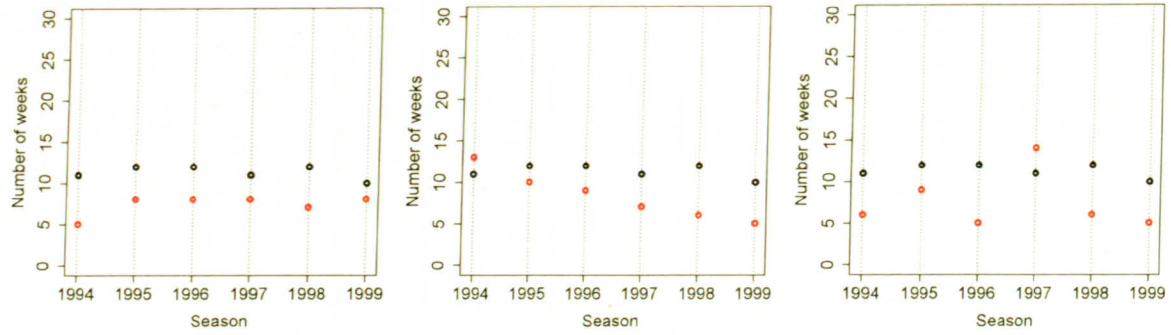
(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64

(e) ILI 0-4 yrs (f) ILI 45-64 (g) ILI ≥ 65

Figure 7.23.: **Comparison of duration of 'aberrant' periods in England & Wales estimated using the method of [1, 2] to define 'aberrant' periods (black circles) and estimated using the HMM (red circles).** Recall that in [1, 2] 'aberrant' periods were defined as weeks when all-age ILI exceeded the upper 95% CI on the average all-age ILI in weeks with <1 all-age laboratory report for influenza.



(a) P&I 0-4 yrs (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64



(e) ILI 0-4 yrs (f) ILI 45-64 (g) ILI ≥ 65

Figure 7.24.: **Comparison of duration of 'aberrant' periods for England & Wales estimated using the method of [3] to define 'aberrant' periods (black circles) and estimated using the HMM (red circles).** Recall that in [3], 'aberrant' periods were defined as weeks surrounding the peak week of all-age laboratory reports encompassing 70% of influenza laboratory reports for that season.

7.12.1. Plausibility of an effect of CTs

Based on a number of pieces of evidence, an average inflating effect of CTs on the mean shift would be expected. First, this association is biologically plausible. CTs would be expected to result in an increase in the proportion of population susceptible to dominant circulating influenza A/H3N2 virus variants since antibodies generated by natural infection or vaccination against previously circulating variants become less able to neutralise variants which are antigenically drifted. Recall from section 2.6.2 that the make-up of the vaccine used in most countries is updated regularly to track the antigenic evolution of the influenza A/H3N2 virus (the vaccine is less regularly updated with a different variant of A/H1N1 virus or lineage of B virus). [106] The recommended influenza A/H3N2 virus variant for the vaccine is updated when there is an antigenic distance of at least 2 units (a fourfold dilution of antiserum in the HI assay) between the vaccine variant and the variant expected to circulate in the next influenza season. The average antigenic distance between clusters of influenza A/H3N2 virus is 4.5 units. There is evidence to suggest that immunity to one variant in a cluster confers protective immunity to challenge with other variants within the same cluster. [103] There appears to be between 60 and 80% cross immunity between clusters adjacent in time. [103, 104] There is probably little or no cross-immunity between non-adjacent clusters. [105]

Second, the positive association between CTs and the mean shift is supported by a modeling study. In section 2.7 it was noted that the best estimate of the average effect of a number of CT seasons on influenza morbidity or mortality is from a model of the antigenic evolution of influenza A/H3N2 virus coupled with a transmission dynamic model. Recall that model-predicted peak rates of 'cases' of influenza during the first season after a new cluster emerged are approximately 1.6 times higher than the average peak rates in other seasons. [14] This value was used to set strongly informative priors on the effect of CTs (and the antigenic distance between clusters) on the mean shift in P&I and ILI in section 7.5.

Third, a positive association between large antigenic drift events and epidemics has often been noted (e.g. in 1972 and 1975 [93]). Most reports in the literature regarding the impact of large antigenic drift events in influenza A/H3N2 virus evolution refer to individual drift events [63, 107, 108]

or to a number of antigenic drift events that coincided with massive epidemics. [58, 93, 109, 110] This means that the average effect of CTs on the relative impact of influenza seasons may be lower than what can be gleaned from the literature. Greene *et al.* plotted the monthly percentage of AC due to P&I from 1968 to 1998 for US residents aged 65 and over; this plot suggested that all CT seasons resulted in an average or above average percentage of deaths due to P&I but that not all of the highest peaks in the graph occurred during CT years. [94]

From work described in chapter 4, crude associations between CTs and peak incidence of ILI or P&I observed in each influenza season appear to be, at most, weakly positive. When peak P&I and ILI rates per season are ranked, only for P&I in the ≥ 65 age group do CT seasons occur in at least five of the top ten seasons. T-tests suggest weak evidence for small increases in peak P&I and ILI in the first H3N2-dominated season after a CT compared with the average season or compared with intracluster seasons. There is no clear association between peak rates of P&I and ILI each influenza season and the antigenic distance between clusters (figure 4.11). Taken together, these points suggest that any effect of CTs or the antigenic distance between clusters on ILI or P&I are small and would be difficult to detect using modeling.

7.12.2. Explanation for what was observed

The magnitude of the mean shift each influenza season is highly variable whether or not a CT has occurred. No strong association between CTs and mean shifts is evident. There is less of a positive association between mean shifts and CTs than between peak incidence each season and CTs. This may be because epidemics are poorly modeled by the HMMs. The inability to fit epidemics dampens variability in mean shifts, making it more difficult to detect the effect of CTs on the variability in mean shifts.

When the mean shifts for each influenza season are ranked largest to smallest, never more than 4 out of 10 of the top seasons (in terms of the size of the mean shift) are CT seasons. This agrees with findings when peak incidence, instead of the mean shift, was ranked in chapter 4. There is no clear association between exponentiated mean shifts and the size of cluster transitions.

There is not sufficient information in the P&I and ILI data analyzed to quantify the magnitude of the effect of CTs, or the antigenic distance between CTs, on the mean shift. This is consistent with the finding that terms for influenza A/H3N2 virus antigenic drift were not important determinants of excess mortality in multiple linear regression analyses. [51, 52] Smaller degrees of antigenic drift were used in these other studies, however, which means these terms would have been less likely to be strong determinants of excess mortality than large antigenic drift events like CTs.

7.12.3. Reasons for a lack of information on CTs effects

There are at least six reasons why the analysis undertaken had limited power to detect an effect of CTs on mean shifts. First, between 1975 and 2004 there were only 9 CTs. This is perhaps too few cluster transitions to allow their average effect to be estimated. The effect of CTs might be heterogeneous. If factors such as vaccine mismatch, the dominant circulating variant(s) (e.g. influenza A/H3N2 virus alone vs. in concert with influenza B or A/H1N1 virus) and ambient temperature modify the effect of cluster transitions, it would be difficult to detect an average effect of CTs on mean shifts, stratified by each of these potential effect modifiers, given only 9 or 10 cluster transitions have been observed. Another possible source of heterogeneity in any CT effect is the overlap in circulation of some clusters. [13] A small overlap in the circulation of two clusters of influenza A/H3N2 viruses may result in the population being sufficiently immunologically primed to blunt the effect of CTs. As a first step in investigating whether the effect of CTs is heterogeneous, the coefficient for the effect of CTs on the mean shift could be modeled as a random effect (see chapter 8).

Second, the effect of a cluster transition may be felt in the first season after it is isolated or in some other season. It may depend on when during the season the transition occurs: a cluster transition occurring early in the season may have its effect felt in the season of isolation whereas a cluster transition occurring later in the season may not produce an effect until a subsequent season. The models fitted only allowed for an effect in the season the CT was first isolated, or in the first H3N2-dominated season after a CT if the season of first isolation was not dominated by influenza A/H3N2 virus. There were convergence problems when models allowing for an effect of CTs

in the season it was isolated as well as in the 1st or 2nd season after it was first isolated were tried. If the effect of CTs occurs sometimes in the season in which it is isolated and sometimes 1 or 2 seasons after it is first isolated, the power to detect the average effect of a CT on the mean shift will be very low given only 9 CTs in the study period.

Third, the biLOGR mean shift may not be a good indicator of the relative impact of influenza seasons. It may be necessary, for example, to take into account the duration of the 'aberrant' period, as well as the mean shift (analogous to analyzing excess morbidity or mortality). In general, higher impact influenza seasons (such as those when influenza A/H3N2 virus is dominant) are also shorter. Taking into account the duration of the 'aberrant' period as well as the mean shift would increase the estimated impact of slow burn and mixed-type influenza seasons (like 1996/97). This would be likely to reduce the variability in the estimated impact of influenza seasons, not increase it, and so would not make it easier to estimate determinants of the variable impact of influenza seasons. Other indicators of the relative impact of influenza seasons have been used historically, for example the percentage of deaths from all-causes that were registered to underlying P&I. [94] This indicator is sensitive to the proportion of deaths due to other causes. It may be fruitful to develop an indicator of the relative impact of influenza seasons in terms of both morbidity and mortality. For example, seasons which result in a large mean shift in both ILI and P&I would be defined as higher impact than seasons resulting in a large mean shift in either ILI or P&I alone. This would downweight seasons with high excess P&I or ILI only, which might be caused by something other than influenza, and consequently increase the specificity of the indicator of the relative impact of influenza seasons for influenza.

Fourth, epidemics are poorly modeled by the HMMs. Because of this, the variability in the impact of influenza seasons may be underestimated. This would make it more difficult to detect the effect of CTs on mean shifts. Fifth, the assumption that CTs affect mean shifts additively on the logarithmic scale may not hold. It was not possible to use multivariate IDR models to estimate covariate effects assuming an additive effect on the original scale.

The sixth reason why the analysis was unable to estimate an effect of CTs is the general computational difficulties of the analysis. This reason is related to the others listed above. Model fit of IDR models was unsatisfac-

tory (recall chapter 6) and so exposure effects were not investigated using IDR models. In IDR models, exposures act additively on the original scale. No model adequately fitted P&I data for the $\geq$ 65 age group. The effect of CTs on deaths could therefore not be estimated (the vast majority of deaths occur in the $\geq$ 65 age group). It may be that the effect of CTs is detectable only in relative mortality between influenza seasons (or mortality + morbidity) but not in morbidity alone.

7.12.4. Plausibility of a vaccine impact

In section 7.9.2, the plausible impact of vaccination on P&I and ILI in the $\geq$ 65 age group in England & Wales was proposed based on vaccine coverage and estimated VE from the literature. For upper estimates of coverage and VE against P&I in the $\geq$ 65 age group, around 55% of excess P&I in this age group might have been prevented through vaccination. It therefore might have been possible to detect an impact of vaccination on the mean shift in P&I had it been possible to model P&I in the elderly (see chapter 6). For upper estimates of coverage and VE against ILI in the elderly, approximately 14% of excess ILI in this age group might have been prevented through vaccination. This estimate is sensitive to the assumptions of constant VE influenza season to influenza season and that vaccinated and unvaccinated elderly are similar to each other in all ways apart from vaccination status as well as to the presumed excess ILI in the absence of vaccination. Since it is likely that children, and not elderly people, are the drivers of transmission of influenza in the community, [38, 39] an impact on P&I or ILI in other age groups due to indirect effects (via herd immunity) of vaccination of the elderly was expected to be nil or too small to detect.

Recall from chapter 4 that there is a weak negative association between vaccine coverage of the $\geq$ 65 age group and peak seasonal P&I and ILI rates across age groups. Estimates of the peak incidence observed in influenza seasons are sensitive to long-term changes to baseline incidence (noted in chapter 4). For this reason the weak negative association between vaccine coverage and peak P&I and ILI is not interpreted as evidence of vaccine impact.

7.12.5. Explanation for what was observed

There is no clear association between increasing vaccine coverage of the elderly and the size of mean shifts in P&I or ILI for any age group.

The impact of vaccination on P&I in the elderly could not be assessed. For models of the impact of vaccine coverage as a quantitative variable, it was not possible to set sensible weakly informative priors (for example, that excluded the possibility that high vaccine coverage is associated with larger mean shifts). This may indicate the assumption of additive vaccine impact on the logarithmic scale is invalid. It was possible to set sensible weakly informative priors on the impact of vaccine coverage as an ordered categorical variable with three levels (no vaccination, moderate coverage and high coverage). There is not sufficient information in the ILI data for the $\geq$ 65 age group to quantify the magnitude of vaccine impact on mean shifts in ILI. Posteriors for coefficients for the impact of moderate (36%) and high (70%) coverage on ILI in the elderly are dominated by the reference or weakly informative priors used.

Two studies, one from the Netherlands and the other from Canada, provide some evidence of vaccine impact of the elderly on influenza-related GP consultations in the elderly, but may overstate impact because of uncontrolled confounding. Dijkstra *et al.* regressed the total ILI rate for the influenza season, not excess ILI rate, on vaccine coverage each influenza season. [121] A decline of 1.7 ILI consultations per 10,000 per % change in coverage was noted (95% CI -3.3 to 0.01). The 95% CI for this result just includes the null. Analyzing total ILI (baseline + excess) is sensitive to changes in the long-term trend in ILI. As such, it would be useful to know whether there have been declines in the use of GP services in the Netherlands as there have been in England & Wales. If so, this decline coincident with increasing vaccination may simply reflect confounding by declines in use of GP services.

Kwong *et al.* estimated RRs comparing excess GP consultations for pneumonia or influenza for two periods: 1997-1999 (before the introduction of a UIIP in Ontario, Canada) and 2000-2003. [95] The authors estimated influenza - attributable outcomes by regressing them on influenza indicators (see section 2.4.4). Looking only at the RRs for the other provinces (not Ontario), the estimated decline in GP consultations for pneumonia or

influenza coincident with the approximately 20% increase in vaccine coverage of the elderly $\geq$ 65 yrs was 85% (95% CI 83 to 87%) in the 65-74 age group, 85% (82 to 87%) in the 75-84 age group and 80% (77 to 83%) in the $\geq$ 85 age group. These estimates of impact were controlled for sex and RSV circulation. Estimated impact was lower when the analysis was restricted to H3N2-dominated seasons. These findings may be confounded by different ambient winter temperatures or different health status of elderly in the period before and after 2000. Viboud and Miller argued that because similar declines were observed in the US as in Ontario (and the other Canadian provinces) comparing the two time periods (2000-2003 to 1997-1999), during which time vaccine coverage of the elderly in the US was relatively stable, an unmeasured factor probably explains a portion of the apparent vaccine impact estimated by Kwong *et al.*. [133]

The impact of vaccination on P&I in the elderly in England & Wales might be large enough to detect. It was not possible to estimate the impact of vaccination on P&I in the elderly using the HMMs. Reports of the impact of vaccination on excess P&I in the elderly in the US and Italy [7,67] do not include upper bounds on CIs for trends so it is not clear whether there was sufficient information in these data to estimate vaccine impact. The authors of both papers report no evidence of a non-zero linear trend coincident with increased vaccine coverage, suggesting that the CIs on linear trend lines were wide and included the null. This suggests that vaccine impact on P&I in these settings may be modest and therefore difficult to detect using linear regression.

Studies from the Netherlands and Canada comparing rates of influenza mortality before and after an increase in vaccine coverage of the elderly suggest that for an increase in coverage of approximately 30% a decline in influenza - attributable mortality of 35-70% could be expected. [64, 95] Despite careful control for confounding by age and RSV circulation, and stratification by high impact/H3N2-dominant seasons, it is unlikely that the declines in mortality noted in the Netherlands and in Canada are entirely attributable to vaccination (see section 2.9.2).

The suggestion of effect modification of vaccine impact in England & Wales by H3N2-dominance found in the analysis undertaken for the thesis is in agreement with evidence for effect modification in other settings. [7, 64,67,95]

7.12.6. Reasons for a lack of information on vaccine impact

Some of the same reasons given for the lack of power to estimate CT effects may explain the inability to detect an impact of vaccination (section 7.12.4). The true vaccine impact on ILI may be too small to detect. The biLOGR mean shift may not be a good indicator of the relative impact of influenza seasons. Epidemics are poorly modeled by HMMs. This means that variability in the relative impact of influenza seasons was underestimated, blunting the ability to estimate determinants of that variability. As noted above, IDR models were abandoned because of poor model fit so the additive on the original scale impact of vaccination as a quantitative variable was not estimated. Modeling vaccine coverage as an ordered categorical variable is computationally easier, though an impact of moderate or high vaccine coverage on ILI in the elderly could not be quantified. Computational difficulties, especially the fact that no model adequately fits P&I data for the $\geq$ 65 age group, severely limited the ability of the analysis to detect an impact of vaccination. Since VE is higher for more severe outcomes, an impact of vaccination on mortality might exist in the absence of an impact on morbidity. It was not possible to assess this in using the HMMs.

7.12.7. Suggested methods for estimating the effect of CTs and vaccine impact on morbidity and mortality

It might be possible to demonstrate impact of yearly vaccination of the elderly by comparing different locations where yearly coverage of the elderly stepped up in different influenza seasons or increased at different rates. Excess morbidity or mortality, or mean shifts in morbidity or mortality, could be compared by influenza season (or number of influenza seasons) across settings where vaccine coverage was very different (e.g. [95]). This would minimise potential confounding by antigenic drift since antigenic drift tends to be geographically homogeneous. [101] Estimates of excess or mean shifts could be controlled for factors which might differ between settings and be associated with excess or mean shifts (like ambient temperature).

Alternatively, a lower bound on the impact of vaccinating elderly people against influenza each season could be demonstrated using the method of Carrat and Valleron and season-specific estimates of VE, excess mortality (or morbidity) and vaccine distribution (see section 2.9.2). [79] For exam-

ple, VE against ILI each season could be estimated from the proportion of ILI patients captured by RCGP WRS who were vaccinated, using the screening method. [217, 218] This value could be substituted into the formula of Carrat and Valleron. Recall that Carrat and Valleron estimated excess mortality by regressing respiratory, cardiovascular and other mortality rates on rates of mortality registered to influenza and an error term that had an ARIMA structure. They then used the estimated excess mortality for each season (d_o), actual vaccine coverage (p) and plausible values of VE to estimate the mortality rate prevented through vaccination (d_a) as $d_a = (d_o VEp)/(1 - VEp)$. This method assumes the attack rate in the unvaccinated is unaffected by vaccination. Therefore using this method would give a lower bound estimate of vaccine impact (because it would encompass only direct effects of vaccinating a proportion of the elderly population).

A method that could be employed to estimate the effect of cluster transitions allowing for confounding by vaccination is a transmission dynamic model. Both large antigenic drift events and increasing vaccine coverage of the elderly is expected to modify the size of the susceptible population (increasing it in the case of antigenic drift, decreasing it with vaccination), specifically the proportion of the population susceptible to the dominant circulating influenza A/H3N2 virus variants. This is not explicitly modeled in the latent variable time series models fitted in the thesis. One way to use a transmission dynamic model to explore the effect of CTs, controlling for vaccine impact, would be to modify the time series-SIRS model of Finkenstadt *et al.*. [88] Recall from section 2.4.6 that the number susceptible (ζ) and recovered (ω) in week t+1 was dependent on a parameter, γ that captured the return of immune individuals to the susceptible class due to waning immunity or antigenic drift

$$\zeta_{t+1} = \zeta_t - \kappa_{t+1} + \gamma_t \omega_t$$
$$\omega_{t+1} = (1 - \gamma_t)\omega_t + \kappa_{t-1}$$

where κ is incidence of ILI determined from a time series of weekly ILI counts and estimates of the proportion of ILI cases who have influenza and the proportion of influenza cases that consult a GP with ILI. To estimate the effect of CTs, γ could be constrained to be constant for the entirety of influenza seasons apart from 1 week during seasons when a CT was ob-

served; this would allow estimation of the effect of CT seasons on population immunity and on predicted ILI. To control for confounding by vaccination, the maximum value for γ, the maximum proportion of recovereds returning to the susceptible class, in a given influenza season could be constrained to be 1-VEp, where p is vaccine coverage. Preliminary work by Finkenstadt *et al.* suggested it would be difficult to apply this method to ILI in England & Wales because the proportion of those with ILI who consult a GP, by age, and the proportion of ILI patients who have influenza are not known for England & Wales. [88] It might be possible to elicit prior information on these factors from experts, like GPs.

8. Discussion and conclusions

The objectives of the work described in the thesis were threefold:

Objective 1: To estimate the relative impact of influenza seasons in England & Wales between 1975/76 and 2004/05 in terms of ILI and P&I by jointly modeling ILI, P&I and laboratory reports for influenza A virus using multivariate latent variable time series models.
Objective 2: To use multivariate latent variable time series models to estimate the mean effect of large antigenic drift events, or cluster transitions, in influenza A/H3N2 virus evolution on the relative impact of influenza seasons in terms of P&I and ILI by age group.
Objective 3: To use the same models to estimate the impact on influenza-attributable ILI and P&I in the $\geq$ 65 age group, and in other age groups, per unit increase in yearly vaccine coverage of the $\geq$ 65 age group.

8.1. Key findings related to objective 1

Excluding high counts to explore long-term trend in P&I and ILI in the absence of influenza virus circulation, it was shown in chapter 4 that there are complex long-term trends in rates of ILI and P&I between 1970 and 2005 which should be flexibly modeled, for example with cubic splines. Trends differ between P&I and ILI and across age groups. There is a decline in the long-term trend in P&I in the $\geq$ 65 age group in England & Wales from approximately 1998 which is coincident with markedly increased vaccine coverage in this age group. Long-term trends in ILI, in the absence of influenza virus circulation, declined in all age groups between the mid 1980s and mid 1990s, were stagnant to 2000, and declined further thereafter. The consistency of the trends in ILI across age groups, and in many

other diagnostic categories, suggests a social or environmental aetiology.

Also in chapter 4, it was highlighted that negative binomial GLMs allowing for seasonality and flexibly modeling trend do not account fully for autocorrelation and overdispersion in ILI and P&I data.

In chapter 5 it was shown that Poisson log-link and identity-link two-state HMMs are sufficient to model P&I and ILI data taking into account autocorrelation and overdispersion. Negative binomial or Gaussian models, which address overdispersion over that captured by the Poisson HMM, do not provide improved model fit or convergence over the Poisson models. The Poisson HMMs are able to distinguish two states ('aberrant' and 'normal') in P&I and ILI data. The relative impact of influenza seasons is captured in a random effect mean shift for each influenza season. The lack of consistency in whether models with an identity link or a log-link are superior for a given data set is unsatisfactory.

In chapter 6, modeling P&I or ILI jointly with laboratory reports for influenza A was shown to increase the precision of the mean shift for most model fits. This is what was expected since jointly modeling P&I or ILI with laboratory reports increases the specificity of the model for influenza. There appears to be a conflict between ILI and laboratory reports for age groups <65 and between P&I and laboratory reports for the ≥ 65 age group in estimating the sequence of 'aberrant' and 'normal' periods in the data. This is because the timing of 'aberrant' periods differs between P&I, ILI and laboratory data. It was shown that the lag that exists between the timing of the 'aberrant' periods in P&I, ILI and laboratory data is not constant between influenza seasons. This is important since it suggests none of P&I, ILI or laboratory data alone would consistently provide the earliest indication that the influenza season had started, year on year. Also, averaging rates of morbidity or mortality in particular weeks across influenza seasons to explore, for example, the average lag between the increase in activity in ILI and the increase in activity in laboratory data would provide a biased estimate of the true lag since it would obscure this variability in the timing of 'aberrant' periods between outcomes across influenza seasons.

8.2. Key findings related to objective 2

In chapter 4 it was noted that, crudely, the distribution of peak rates for CT seasons appears greater than for intracluster seasons. Ranking of peak P&I and ILI rates observed across seasons revealed that only for P&I in the ≥ 65 age group do CT seasons occur in at least five of the top ten seasons. T-tests suggest weak evidence for small increases in peak P&I and ILI (6 P&I per 1,000,000 population, 95 ILI per 100,000 population) in CT seasons compared with the average season. There is weak evidence for 8/1,000,000 high rates of P&I and 133/100,000 higher rates of ILI in CT compared with intracluster seasons. Such small differences are of little public health importance given that, in the data analyzed in the thesis, weekly rates of P&I of up to 80/1,000,000 and ILI of up to 2,322/100,000 are observed. It was also noted in chapter 4 that there is no clear association between peak rates of P&I and ILI each influenza season and the antigenic distance between clusters.

In chapter 7, scatter plots of mean shifts against covariates show little evidence of association between CTs, or the antigenic distance between clusters, and mean shifts in P&I or ILI. biLOGR HMMs fitted with a dependency between the mean shift and CTs, or the antigenic distance between clusters, do not provide evidence of an average inflating effect of CTs on the relative impact of influenza seasons. The difficulty that the models had in sampling from reference prior models, and the dominance of informative priors on posterior distributions, suggests there is little information in the data with which to quantify the effect of CTs on the mean shift in P&I or ILI.

8.3. Key findings related to objective 3

In chapter 4 it was shown that there is a weak negative association between vaccine coverage of the ≥ 65 age group and peak seasonal P&I and ILI rates across age groups. In chapter 7, scatter plots of mean shifts in P&I and ILI, by age group, against vaccine coverage of the ≥ 65 age group suggested little evidence of an association. Stratifying mean shifts by the dominant variant in the influenza season (influenza A/H3N2 vs H1N1/B) showed that H3N2-dominance may act as an effect modifier of vaccine impact. This is

because VE would be higher in higher impact influenza seasons during which influenza A/H3N2 virus tends to be dominant. There is not a consistent association between vaccine coverage and mean shifts in H3N2-dominated seasons. There is no association between vaccine coverage and mean shifts in seasons dominated by influenza A/H1N1 or B virus. It is unlikely that influenza A/H3N2 virus-dominance is acting as a confounder of vaccine impact since influenza A/H3N2 virus-dominated seasons are not more or less frequent during the period when vaccine coverage was moderate or when coverage was high.

From attempts to quantify vaccine impact from model fits, also in chapter 7, it was shown that there is limited information in the data about vaccine impact. Posteriors for the coefficient for vaccine impact are little influenced by the data and are dominated by priors.

8.4. Strengths

The work undertaken for the thesis has a number of strengths. First, the modeling of long-term trend in P&I and ILI, in the absence of influenza, with cubic splines has increased the flexibility of fitting seasonality with the sine and cosine term. Careful control for long-term changes in morbidity and mortality has been achieved. Second, the two-state Poisson HMMs have adequately modeled variability in the P&I and ILI data meaning that account has been taken for overdispersion of these data relative to the Poisson distribution. Third, by fitting two-state HMMs, where the model itself determines whether the data are consistent with two states, there is no need to designate 'aberrant' from 'normal' incidence externally to model fitting. Fourth, fitting multivariate HMMs, where the designation of 'aberrant' from 'normal' incidence was made jointly by information in the P&I or ILI data as well as in laboratory reports, increases the specificity of the model-estimated 'aberrant' periods in P&I and ILI for influenza. BiLOGR models fitted to either P&I or ILI allow for different 'aberrant' periods in the two data sets, for example due to the deaths being lagged relative to consultations.

8.5. Limitations

There are also limitations of the work described in the thesis, both related to the data and to the modeling.

8.5.1. Limitations of the data

When do CTs become dominant?

The seasons when cluster transitions were first identified in the WHO vaccine strain selection data set were taken from the paper by Smith *et al.*. [13] Recall that much of the information in the data set, on influenza variants and on what dates they were isolated from patients, was from the Netherlands. It is not known when during the particular influenza season the cluster transition variant began to circulate in England & Wales. The season of first isolation of the CT, or the first H3N2-dominated season after this, might not be a good proxy for when the new cluster became dominant in England & Wales. There is uncertainty as to whether the new cluster would have been dominant for the duration of the season of its emergence or for only, say, the final month of that first season (because it only became the dominant circulating H3N2 variant one month before the end of the season). There is also some overlap in the circulation of adjacent clusters, adding to the uncertainty in when a particular cluster was dominant. It is possible that variability in the timing of dominance of clusters may partly explain the difficulty in detecting an effect of CTs on the mean shift.

A related issue is that the dominant influenza variants circulating in each influenza season in England & Wales were taken from the literature (table 3.2). The percentage of subtyped isolates that were influenza A/H3N2 virus or influenza A/H1N1 virus was not available for most influenza seasons. Therefore, in addition to the uncertainty about when exactly each cluster became dominant in England & Wales, there is also uncertainty about what proportion of circulating variants in any season were influenza A/H3N2 virus variants.

Dates

The time-varying lag between the timing of 'aberrant' periods across outcome variables, identified in chapter 6, is likely to be partly biological and

partly a function of the data, especially the laboratory data. One possible biological reason for different lags between outcome variables across influenza seasons may be variability in the virulence of circulating variants in different influenza seasons. More virulent variants may result in mortality sooner after their initial appearance than less virulent variants. Other possible biological explanations include the different types and subtypes of influenza circulating in different influenza seasons or interference in some influenza seasons by cocirculating viruses. In laboratory data, the change in 1993 to include in LabBase2 the positive reports from one of two new sentinel swabbing studies of ILI patients [33] means that the relationship between the non-laboratory confirmed outcome variables (P&I and ILI) and laboratory data changed at that point. Before 1993 most laboratory testing was done in hospitals; from 1993 onwards the data in LabBase2 are a mixture of hospital and sentinel GP laboratory reports. The lag between timing of 'aberrant' periods in P&I and ILI did not appear to become time-varying around 1993, or to have had one value before 1993 and another value after 1993, so this artefact probably does not fully explain the time varying lag. The time varying lag may also be partly due simply to random variation.

Hospitalisations

Part of the morbidity burden associated with influenza includes respiratory hospitalisations. [49, 76] Vaccine impact, or the effect of CTs, on rates of respiratory hospitalisation could be substantial and has not be addressed in the thesis. However, the maximum number of hospital beds may blunt estimates of the relative impact of influenza seasons in terms of hospitalisation. [42] This would limit their usefulness in estimating the relative impact of influenza seasons in terms of morbidity and determinants of this relative impact.

8.5.2. Limitations of the modeling

Autocorrelation plots of residuals show there is inadequate modeling of seasonality in two-state HMMs. Future work could include modeling seasonality more flexibly, e.g. with splines with several df per year [42] or monthly indicators, [184] instead of with the sine and cosine terms.

There is also inadequate modeling of large peaks in incidence. This is

partly due to the mean shift being the average ratio of incidence in the 'aberrant' to 'normal' weeks for a given influenza season and likely also due in part to the absence of important explanatory variables from the models. In addition, P&I and ILI data may be consistent with more than two states. HMMs with two states were fitted for the thesis since a biological explanation can be assigned to two states (i.e. incidence in the absence of influenza from one state and incidence attributable to influenza from a second state). Future work could explore fitting models with more than two states in an attempt to better model epidemics. For example, three states could be conceptualised as 'normal', 'aberrant (but not epidemic)' and 'epidemic' incidence.

There is a time-varying lag between state transitions for outcome variables within an age group (<65 for ILI and laboratory data and ≥ 65 for P&I and laboratory data) and this is not allowed for in any model.

Multivariate state sequences include some false alarms, where weeks are labeled as 'aberrant' though the high incidence is probably unrelated to influenza. For example, several models define a short summer period as 'aberrant' in some years, perhaps because of heat waves. An area for further development within the HMM framework would be to weight outcome variables for state sequence estimation (for example, more heavily weighting laboratory vs. other outcomes) to reduce the probability of false alarms in the state sequence. This would increase the specificity of the mean shift for influenza. The most specific state sequence for influenza would result from estimating the state sequence solely using the laboratory data. The state sequence estimated using only laboratory data could then be applied to the P&I and ILI data for estimation of mean shifts. This would produce the most specific mean shifts for influenza but would misclassify some influenza - attributable morbidity or mortality as 'normal' because of lags between influenza-attributable incidence in laboratory and non-laboratory data and the fact that laboratory and non-laboratory data do not represent the same types of people in terms of comorbid illness, for example (see section 3.2.6).

In relation to attempts to use the HMMs to estimate effects of exposures of interest, the models do not allow for heterogeneity in covariate effects. Heterogeneity in the effect of CTs on the mean shift could be allowed for by modeling the coefficient for CTs as a random effect. There may be spatial heterogeneity in vaccine coverage or VE, if intensity of influenza

transmission is greater in northern than southern areas of England & Wales. Models allowing for spatial heterogeneity in relative impact of influenza seasons and in covariate effects may therefore provide increased power to estimate vaccine impact.

Some of the influenza - attributable P&I and ILI incidence may have been modeled by the cubic splines. Overfitting of the long-term trend could have occurred because the number of df for the splines was decided based on an analysis of data from 1970-2005 (chapter 4) whereas HMMs were fitted to data from 1975/76-2004/05.

It is unclear why no multivariate HMM fitted to P&I in the $\geq$ 65 age group converged. Even the univariate LOGR HMM fitted to P&I for the $\geq$ 65 age group failed to converge. The whole explanation for this cannot be a conflict between P&I and laboratory data because in the LOGR model, only P&I data were modeled. For the HMM framework to be most useful for modeling influenza-related data in England & Wales, it is necessary to find a way to use the HMM to model deaths in the elderly. Risk of death increases with older age. Analyzing data in the elderly in finer age bands may therefore also provide additional information with which to estimate covariate effects.

Some of the problems of convergence of HMMs undoubtedly relate to sensitivity to model assumptions and not only to limited information in the data about the state sequence. Recall that in chapter 5 negative binomial and Gaussian HMMs were discarded in favour of Poisson HMMs because of lack of convergence. A possible explanation for the problems of the negative binomial and Gaussian HMMs is that these models fail to distinguish overdispersion relative to the Poisson distribution from variability generated by the hidden Markov process. The reliance on Poisson HMMs in the remainder of the thesis, where variability is generated solely by a mixture of Poisson distributions, is therefore probably optimistic. The variability generated by the mixture of Poisson distributions may inadequately address the various potential explanations for overdispersion (like variation in reporting and spatiotemporal clustering of cases). Incorporating some structure into the model, where the Markov chain models the transmission of infection from infected to susceptible people, might help the model to distinguish the underlying state process from clustering of cases due to the infectious nature of influenza [89] and thus lead to improved convergence. Detailed

spatial data, if available, could be used to allow reconstruction of chains of transmission by structuring the model in both space and time. This further model development would probably necessitate abandoning OpenBUGS in favour of other packages that might more efficiently handle the increased model complexity.

Bibliography

[1] D. M. Fleming, "The contribution of influenza to combined acute respiratory infections, hospital admissions, and deaths in winter," *Commun Dis Public Health*, vol. 3, no. 1, pp. 32–8, 2000.

[2] D. M. Fleming, R. S. Pannell, A. J. Elliot, and K. W. Cross, "Respiratory illness associated with influenza and respiratory syncytial virus infection," *Arch Dis Child*, vol. 90, no. 7, pp. 741–6, 2005.

[3] D. M. Fleming, A. J. Elliot, and K. W. Cross, "Morbidity profiles of patients consulting during influenza and respiratory syncytial virus active periods," *Epidemiol Infect*, pp. 1–10, 2007.

[4] D. Fleming and M. Wood, "The clinical diagnosis of influenza," *Curr Med Res Opin*, vol. 18, no. 6, pp. 338–41, 2002.

[5] C. E. Long, C. B. Hall, C. K. Cunningham, L. B. Weiner, K. P. Alger, M. Gouveia, C. B. Colella, K. C. Schnabel, and W. H. Barker, "Influenza surveillance in community-dwelling elderly compared with children," *Arch Fam Med*, vol. 6, no. 5, pp. 459–65, 1997.

[6] D. M. Fleming, M. Zambon, A. I. Bartelds, and J. C. de Jong, "The duration and magnitude of influenza epidemics: a study of surveillance data from sentinel general practices in England, Wales and the Netherlands," *Eur J Epidemiol*, vol. 15, no. 5, pp. 467–73, 1999.

[7] L. Simonsen, T. A. Reichert, C. Viboud, W. C. Blackwelder, R. J. Taylor, and M. A. Miller, "Impact of influenza vaccination on seasonal mortality in the US elderly population," *Arch Intern Med*, vol. 165, no. 3, pp. 265–72, 2005.

[8] W. W. Thompson, D. K. Shay, E. Weintraub, L. Brammer, N. Cox, L. J. Anderson, and K. Fukuda, "Mortality associated with influenza

and respiratory syncytial virus in the United States," *JAMA*, vol. 289, no. 2, pp. 179–186, 2003.

[9] R. J. Pitman, A. Melegaro, D. Gelb, M. R. Siddiqui, N. J. Gay, and W. J. Edmunds, "Assessing the burden of influenza and other respiratory infections in England and Wales," *J Infect*, vol. 54, no. 6, pp. 530–538, 2007.

[10] A. D. Grant and B. Eke, "Application of information technology to the laboratory reporting of communicable disease in England and Wales," *Commun Dis Rep CDR Rev*, vol. 3, no. 6, pp. R75–8, 1993.

[11] W. M. Fitch, R. M. Bush, C. A. Bender, and N. J. Cox, "Long term trends in the evolution of H(3) HA1 human influenza type A," *Proc Natl Acad Sci U S A*, vol. 94, no. 15, pp. 7712–8, 1997.

[12] N. M. Ferguson, A. P. Galvani, and R. M. Bush, "Ecological and immunological determinants of influenza evolution," *Nature*, vol. 422, no. 6930, pp. 428–33, 2003.

[13] D. J. Smith, A. S. Lapedes, J. C. de Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. M. E. Osterhaus, and R. A. M. Fouchier, "Mapping the antigenic and genetic evolution of influenza virus," *Science*, vol. 305, no. 5682, pp. 371–376, 2004.

[14] K. Koelle, S. Cobey, B. Grenfell, and M. Pascual, "Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans," *Science*, vol. 314, no. 5807, pp. 1898–903, 2006.

[15] T. M. Govaert, C. T. Thijs, N. Masurel, M. J. Sprenger, G. J. Dinant, and J. A. Knottnerus, "The efficacy of influenza vaccination in elderly individuals: A randomized double-blind placebo-controlled trial," *JAMA*, vol. 272, no. 21, pp. 1661–5, 1994.

[16] T. O. Jefferson, D. Rivetti, C. Di Pietrantonj, A. Rivetti, and V. Demicheli, "Vaccines for preventing influenza in healthy adults," *Cochrane Database Syst Rev*, no. 2, p. CD001269, 2007.

[17] S. E. Ohmit, J. C. Victor, E. R. Teich, R. K. Truscon, J. R. Rotthoff, D. W. Newton, S. A. Campbell, M. L. Boulton, and A. S. Monto,

"Prevention of symptomatic seasonal influenza in 2005-2006 by inactivated and live attenuated vaccines," *J Infect Dis*, vol. 198, no. 3, pp. 312–317, 2008.

[18] K. G. Nicholson, J. M. Wood, and M. Zambon, "Influenza," *Lancet*, vol. 362, no. 9397, pp. 1733–1745, 2003.

[19] S. Mubareka, A. C. Lowen, J. Steel, A. L. Coates, A. Garcia-Sastre, and P. Palese, "Transmission of influenza virus via aerosols and fomites in the guinea pig model," *J Infect Dis*, vol. 199, no. 6, pp. 858–865, 2009.

[20] R. G. Webster, W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka, "Evolution and ecology of influenza A viruses," *Microbiol Rev*, vol. 56, no. 1, pp. 152–79, 1992.

[21] M. C. Zambon, "Epidemiology and pathogenesis of influenza," *J Antimicrob Chemother*, vol. 44 Suppl B, pp. 3–9, 1999.

[22] B. D. Greenbaum, A. J. Levine, G. Bhanot, and R. Rabadan, "Patterns of evolution and host gene mimicry in influenza and other RNA viruses," *PLoS Pathog*, vol. 4, no. 6, p. e1000079, 2008.

[23] L. Simonsen, M. J. Clarke, L. B. Schonberger, N. H. Arden, N. J. Cox, and K. Fukuda, "Pandemic versus epidemic influenza mortality: a pattern of changing age distribution," *J Infect Dis*, vol. 178, no. 1, pp. 53–60, 1998.

[24] T. J. Francis, "On the doctrine of original antigenic sin," *Proc Am Philos Soc*, vol. 104, no. 6, pp. 572–578, 1960.

[25] S. C. Schoenbaum, M. T. Coleman, W. R. Dowdle, and S. R. Mostow, "Epidemiology of influenza in the elderly: evidence of virus recycling," *Am J Epidemiol*, vol. 103, no. 2, pp. 166–73, 1976.

[26] E. C. Claas, A. D. Osterhaus, R. van Beek, J. C. De Jong, G. F. Rimmelzwaan, D. A. Senne, S. Krauss, K. F. Shortridge, and R. G. Webster, "Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus," *Lancet*, vol. 351, no. 9101, pp. 472–7, 1998.

[27] G. J. D. Smith, D. Vijaykrishna, J. Bahl, S. J. Lycett, M. Worobey, O. G. Pybus, S. K. Ma, C. L. Cheung, J. Raghwani, S. Bhatt, J. S. M. Peiris, Y. Guan, and A. Rambaut, "Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic," *Nature*, vol. 459, no. 7250, pp. 1122–1125, 2009.

[28] "Serum cross-reactive antibody response to a novel influenza A (H1N1) virus after vaccination with seasonal influenza vaccine," *MMWR CDC Surveill Summ*, vol. 58, no. 19, pp. 521–524, 2009.

[29] R. J. Garten, C. T. Davis, C. A. Russell, B. Shu, S. Lindstrom, A. Balish, W. M. Sessions, X. Xu, E. Skepner, V. Deyde, M. Okomo-Adhiambo, L. Gubareva, J. Barnes, C. B. Smith, S. L. Emery, M. J. Hillman, P. Rivailler, J. Smagala, M. de Graaf, D. F. Burke, R. A. M. Fouchier, C. Pappas, C. M. Alpuche-Aranda, H. Lopez-Gatell, H. Olivera, I. Lopez, C. A. Myers, D. Faix, P. J. Blair, C. Yu, K. M. Keene, J. Dotson, P. David, D. Boxrud, A. R. Sambol, S. H. Abid, K. St. George, T. Bannerman, A. L. Moore, D. J. Stringer, P. Blevins, G. J. Demmler-Harrison, M. Ginsberg, P. Kriner, S. Waterman, S. Smole, H. F. Guevara, E. A. Belongia, P. A. Clark, S. T. Beatrice, R. Donis, J. Katz, L. Finelli, C. B. Bridges, M. Shaw, D. B. Jernigan, T. M. Uyeki, D. J. Smith, A. I. Klimov, and N. J. Cox, "Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans," *Science*, p. 1176225, 2009.

[30] A. P. Kendal, G. R. Noble, J. J. Skehel, and W. R. Dowdle, "Antigenic Similarity of Influenza-A (H1N1) Viruses from Epidemics in 1977-1978 to Scandinavian Strains Isolated in Epidemics of 1950-1951," *Virology*, vol. 89, no. 2, pp. 632–636, 1978.

[31] K. Nakajima, U. Desselberger, and P. Palese, "Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950," *Nature*, vol. 274, no. 5669, pp. 334–339, 1978.

[32] V. Gregory, M. Bennett, M. H. Orkhan, S. Al Hajjar, N. Varsano, E. Mendelson, M. Zambon, J. Ellis, A. Hay, and Y. P. Lin, "Emergence of influenza A H1N2 reassortant viruses in the human population during 2001," *Virology*, vol. 300, no. 1, pp. 1–7, 2002.

[33] C. A. Joseph, "Virological surveillance of influenza in England and Wales: results of a two year pilot study 1993/94 and 1994/95. PHLS Influenza Subcommittee," *Commun Dis Rep CDR Rev*, vol. 5, no. 10, pp. R141–5, 1995.

[34] D. M. Fleming, P. Chakraverty, C. Sadler, and P. Litton, "Combined clinical and virological surveillance of influenza in winters of 1992 and 1993-4," *BMJ*, vol. 311, no. 7000, pp. 290–291, 1995.

[35] R. E. Hope-Simpson, "The role of season in the epidemiology of influenza," *J Hyg (Lond)*, vol. 86, no. 1, pp. 35–47, 1981.

[36] M. C. Zambon, J. D. Stockton, J. P. Clewley, and D. M. Fleming, "Contribution of influenza and respiratory syncytial virus to community cases of influenza-like illness: an observational study," *Lancet*, vol. 358, no. 9291, pp. 1410–1416, 2001.

[37] D. Turner, A. Wailoo, K. Nicholson, N. Cooper, A. Sutton, and K. R. Abrams, "Systematic review and economic decision modelling for the prevention and treatment of influenza A and B," tech. rep., Health Technology Assessment, 2003.

[38] J. P. Fox, C. E. Hall, M. K. Cooney, and H. M. Foy, "Influenzavirus infections in Seattle families, 1975-1979. I. Study design, methods and the occurrence of infections by time and age," *Am J Epidemiol*, vol. 116, no. 2, pp. 212–27, 1982.

[39] A. S. Monto and F. Kioumehr, "The Tecumseh study of respiratory illness. IX. Occurence of influenza in the community, 1966–1971," *Am J Epidemiol*, vol. 102, no. 6, pp. 553–63, 1975.

[40] A. S. Monto, J. S. Koopman, and I. M. J. Longini, "Tecumseh study of illness. XIII. Influenza infection and disease, 1976-1981," *Am J Epidemiol*, vol. 121, no. 6, pp. 811–22, 1985.

[41] M. Curwen, K. Dunnell, and J. Ashley, "Hidden influenza deaths: 1989-90.," *Popul Trends*, vol. 61, no. 31-33., 1990.

[42] P. Mangtani, S. Hajat, S. Kovats, P. Wilkinson, and B. Armstrong, "The association of respiratory syncytial virus infection and influenza

with emergency admissions for respiratory disease in London: an analysis of routine surveillance data," *Clin Infect Dis*, vol. 42, no. 5, pp. 640–6, 2006.

[43] W. W. Thompson, D. K. Shay, E. Weintraub, L. Brammer, C. B. Bridges, N. J. Cox, and K. Fukuda, "Influenza-associated hospitalizations in the United States," *JAMA*, vol. 292, no. 11, pp. 1333–40, 2004.

[44] C. R. Meier, P. N. Napalkov, Y. Wegmuller, T. Jefferson, and H. Jick, "Population-based study on incidence, risk factors, clinical complications and drug utilisation associated with influenza in the United Kingdom," *Eur J Clin Microbiol Infect Dis*, vol. 19, no. 11, pp. 834–42, 2000.

[45] W. H. Barker and J. P. Mullooly, "Underestimation of the role of pneumonia and influenza in causing excess mortality.," *Am J Public Health*, vol. 71, pp. 643–645, Jun 1981.

[46] D. L. Schanzer, J. M. Langley, and T. W. Tam, "Co-morbidities associated with influenza-attributed mortality, 1994-2000, Canada," *Vaccine*, vol. 26, no. 36, pp. 4697–703, 2008.

[47] M. Curwen and T. Devis, "Winter mortality, temperature and influenza: has the relationship changed in recent years?," *Popul Trends*, vol. 54, pp. 17–20, 1988.

[48] L. Simonsen, K. Fukuda, L. B. Schonberger, and N. J. Cox, "The impact of influenza epidemics on hospitalizations," *J Infect Dis*, vol. 181, no. 3, pp. 831–837, 2000.

[49] J. S. Nguyen-Van-Tam, C. R. Brockway, J. C. Pearson, A. C. Hayward, and D. M. Fleming, "Excess hospital admissions for pneumonia and influenza in persons >or = 65 years associated with influenza epidemics in three English health districts: 1987-95," *Epidemiol Infect*, vol. 126, no. 1, pp. 71–9, 2001.

[50] D. Fleming, S. Harcourt, and G. Smith, "Influenza and adult hospital admissions for respiratory conditions in England 1989-2001," *Commun Dis Public Health*, vol. 6, no. 3, pp. 231–7, 2003.

[51] H. E. Tillett, J. W. Smith, and C. D. Gooch, "Excess deaths attributable to influenza in England and Wales: age at death and certified cause," *Int J Epidemiol*, vol. 12, no. 3, pp. 344–52, 1983.

[52] R. E. Clifford, J. W. Smith, H. E. Tillett, and P. J. Wherry, "Excess mortality associated with influenza in England and Wales," *Int J Epidemiol*, vol. 6, no. 2, pp. 115–28, 1977.

[53] D. M. Fleming, K. W. Cross, and R. S. Pannell, "Influenza and its relationship to circulatory disorders," *Epidemiol Infect*, vol. 133, no. 2, pp. 255–62, 2005.

[54] K. Choi and S. B. Thacker, "Mortality during influenza epidemics in the United States, 1967-1978," *Am J Public Health*, vol. 72, no. 11, pp. 1280–3, 1982.

[55] A. S. Monto, S. E. Ohmit, J. R. Margulies, and A. Talsma, "Medical practice-based influenza surveillance: viral prevalence and assessment of morbidity," *Am J Epidemiol*, vol. 141, no. 6, pp. 502–6, 1995.

[56] L. Simonsen, M. J. Clarke, D. F. Stroup, G. D. Williamson, N. H. Arden, and N. J. Cox, "A method for timely assessment of influenza-associated mortality in the United States," *Epidemiology*, vol. 8, no. 4, pp. 390–5, 1997.

[57] D. M. Fleming, R. S. Pannell, and K. W. Cross, "Mortality in children from influenza and respiratory syncytial virus," *J Epidemiol Community Health*, vol. 59, no. 7, pp. 586–90, 2005.

[58] F. Assaad, W. C. Cockburn, and T. K. Sundaresan, "Use of excess mortality from respiratory diseases in the study of influenza," *Bull World Health Organ*, vol. 49, no. 3, pp. 219–33, 1973.

[59] K. G. Nicholson, "Impact of influenza and respiratory syncytial virus on mortality in England and Wales from January 1975 to December 1990," *Epidemiol Infect*, vol. 116, no. 1, pp. 51–63, 1996.

[60] W. W. Thompson, E. Weintraub, P. Dhankhar, P. Y. Cheng, L. Brammer, M. I. Meltzer, J. S. Bresee, and D. K. Shay, "Estimates of US influenza-associated deaths made using four different methods," *Influenza Other Respi Viruses*, vol. 3, no. 1, pp. 37–49, 2009.

[61] C. Viboud, P. Y. Boelle, K. Pakdaman, F. Carrat, A. J. Valleron, and A. Flahault, "Influenza epidemics in the United States, France, and Australia, 1972-1997," *Emerg Infect Dis*, vol. 10, no. 1, pp. 32–9, 2004.

[62] D. L. Schanzer, T. W. Tam, J. M. Langley, and B. T. Winchester, "Influenza-attributable deaths, Canada 1990-1999," *Epidemiol Infect*, vol. 135, no. 7, pp. 1109–16, 2007.

[63] N. Sugaya and Y. Takeuchi, "Mass vaccination of schoolchildren against influenza and its impact on the influenza-associated mortality rate among children in Japan," *Clin Infect Dis*, vol. 41, no. 7, pp. 939–47, 2005.

[64] A. G. Jansen, E. A. Sanders, K. L. Nichol, A. M. van Loon, A. W. Hoes, and E. Hak, "Decline in influenza-associated mortality among Dutch elderly following the introduction of a nationwide vaccination program," *Vaccine*, vol. 26, no. 44, pp. 5567–74, 2008.

[65] L. Denoeud, C. Turbelin, S. Ansart, A. J. Valleron, A. Flahault, and F. Carrat, "Predicting pneumonia and influenza mortality from morbidity data," *PLoS ONE*, vol. 2, no. 5, p. e464, 2007.

[66] P. Zucs, U. Buchholz, W. Haas, and H. Uphoff, "Influenza associated excess mortality in Germany, 1985-2001," *Emerg Themes Epidemiol*, vol. 2, p. 6, 2005.

[67] C. Rizzo, C. Viboud, E. Montomoli, L. Simonsen, and M. A. Miller, "Influenza-related mortality in the Italian elderly: no decline associated with increasing vaccination coverage," *Vaccine*, vol. 24, no. 42-43, pp. 6468–75, 2006.

[68] W. Farr, *Vital Statistics*. London: Office of the Sanitary Institute, 1885.

[69] S. Collins, "Excess mortality from causes other than influenza and pneumonia during influenza epidemics," *Public Health Rep*, vol. 47, no. 46, pp. 2159–2180, 1932.

[70] R. E. Serfling, "Methods of current statistical analysis of excess pneumonia-influenza death.," *Public Health Rep*, vol. 78, no. 6, pp. 494–506., 1963.

[71] Influenza/Respiratory Virus Team, "HPA national influenza season summary: Summary of UK surveillance of influenza and other seasonal respiratory illness 2007/08," tech. rep., HPA Centre for Infections, 2008.

[72] "Flu activity & surveillance: reports & surveillance methods in the United States." http://www.cdc.gov/flu/weekly/fluactivity.htm, Accessed 30 October 2007.

[73] K. J. Lui and A. P. Kendal, "Impact of influenza epidemics on mortality in the United States from October 1972 to May 1985," *Am J Public Health*, vol. 77, no. 6, pp. 712–6, 1987.

[74] K. Choi and S. B. Thacker, "An evaluation of influenza mortality surveillance, 1962-1979. I. Time series forecasts of expected pneumonia and influenza deaths," *Am J Epidemiol*, vol. 113, no. 3, pp. 215–26, 1981.

[75] G. Box and G. Jenkins, *Time-series analysis, forecasting and control.* San Francisco: Holden-Day, revised edition 1976 ed., 1970.

[76] H. S. Izurieta, W. W. Thompson, P. Kramarz, D. K. Shay, R. L. Davis, F. DeStefano, S. Black, H. Shinefield, and K. Fukuda, "Influenza and the rates of hospitalization for respiratory disease among infants and young children," *N Engl J Med*, vol. 342, no. 4, pp. 232–9, 2000.

[77] B. G. Armstrong, P. Mangtani, A. Fletcher, S. Kovats, A. McMichael, S. Pattenden, and P. Wilkinson, "Effect of influenza vaccination on excess deaths occurring during periods of high circulation of influenza: cohort study in elderly people," *BMJ*, vol. 329, no. 7467, p. 660, 2004.

[78] D. L. Schanzer, J. M. Langley, and T. W. Tam, "Hospitalization attributable to influenza and other viral respiratory illnesses in Canadian children," *Pediatr Infect Dis J*, vol. 25, no. 9, pp. 795–800, 2006.

[79] F. Carrat and A. J. Valleron, "Influenza mortality among the elderly in France, 1980-90: how many deaths may have been avoided through

vaccination?," *J Epidemiol Community Health*, vol. 49, no. 4, pp. 419–25, 1995.

[80] N. J. Gay, N. J. Andrews, C. L. Trotter, and W. J. Edmunds, "Estimating deaths due to influenza and respiratory syncytial virus," *JAMA*, vol. 289, no. 19, pp. 2499; author reply 2500–2, 2003.

[81] Y. Le Strat and F. Carrat, "Monitoring epidemiologic surveillance data using hidden Markov models," *Stat Med*, vol. 18, no. 24, pp. 3463–78, 1999.

[82] A. Ozonoff, S. Sukpraprut, and P. Sebastiani, "Modeling seasonality of influenza with hidden Markov models." Joint Statistical Meeting, 2006. Seattle, Washington.

[83] L. Held, M. Hofmann, M. Hohle, and V. Schmid, "A two-component model for counts of infectious diseases," *Biostatistics*, vol. 7, no. 3, pp. 422–37, 2006.

[84] D. Conesa, A. Lopez-Quillez, and M.-A. Martinez-Beneito, "A Bayesian methodology for on-line detecting the onset of influenza epidemics," in *International Biometric Conference*, (Dublin), 2008.

[85] M. Martinez-Beneito, D. Conesa, A. Lopez-Quilez, and A. Lopez-Maside, "Bayesian Markov switching models for the early detection of influenza epidemics," *Stat Med*, vol. 27, no. 22, pp. 4455–4468, 2008.

[86] T. Rath, M. Carreras, and P. Sebastiani, "Automated detection of influenza epidemics with hidden Markov models," in *Cryptographic Hardware and Embedded Systems - CHES 2003*, pp. 521–532, 2003.

[87] B. F. Finkenstadt and B. T. Grenfell, "Time series modelling of childhood diseases: a dynamical systems approach," *J R Stat Soc C*, vol. 49, no. 2, pp. 187–205, 2000.

[88] B. F. Finkenstadt, A. Morton, and D. A. Rand, "Modelling antigenic drift in weekly flu incidence," *Stat Med*, vol. 24, no. 22, pp. 3447–61, 2005.

[89] B. Cooper and M. Lipsitch, "The analysis of hospital infection data using hidden Markov models," *Biostatistics*, vol. 5, no. 2, pp. 223–37, 2004.

[90] P. F. Wright, J. Thompson, and D. T. Karzon, "Differing virulence of H1N1 and H3N2 influenza strains," *Am J Epidemiol*, vol. 112, no. 6, pp. 814–9, 1980.

[91] A. L. Frank, L. H. Taber, and J. M. Wells, "Comparison of infection rates and severity of illness for influenza A subtypes H1N1 and H3N2," *J Infect Dis*, vol. 151, no. 1, pp. 73–80, 1985.

[92] L. Simonsen, M. J. Clarke, G. D. Williamson, D. F. Stroup, N. H. Arden, and L. B. Schonberger, "The impact of influenza epidemics on mortality: introducing a severity index," *Am J Public Health*, vol. 87, no. 12, pp. 1944–1950, 1997.

[93] D. C. Wiley, I. A. Wilson, and J. J. Skehel, "Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation," *Nature*, vol. 289, no. 5796, pp. 373–8, 1981.

[94] S. K. Greene, E. L. Ionides, and M. L. Wilson, "Patterns of influenza-associated mortality among US elderly by geographic region and virus subtype, 1968-1998," *Am J Epidemiol*, vol. 163, no. 4, pp. 316–326, 2006.

[95] J. C. Kwong, T. A. Stukel, J. Lim, A. J. McGeer, R. E. G. Upshur, H. Johansen, C. Sambell, W. W. Thompson, D. Thiruchelvam, F. Marra, L. W. Svenson, and D. G. Manuel, "The effect of universal influenza immunization on mortality and health care use," *PLoS Med*, vol. 5, no. 10, p. e211, 2008.

[96] M. Davy, "Time and generational trends in smoking among men and women in Great Britain, 1972-2004/05," *Health Stat Quarterly*, vol. 32, pp. 35–43, 2006.

[97] A. J. Elliot, K. W. Cross, and D. M. Fleming, "Acute respiratory infections and winter pressures on hospital admissions in England and Wales 1990-2005," *J Public Health*, p. fdn003, 2008.

[98] R. E. Owen, E. Yamada, C. I. Thompson, L. J. Phillipson, C. Thompson, E. Taylor, M. Zambon, H. M. I. Osborn, W. S. Barclay, and P. Borrow, "Alterations in receptor binding properties of recent human influenza H3N2 viruses are associated with reduced natural killer cell lysis of infected cells," *J Virol*, vol. 81, pp. 11170–11178, 2007.

[99] C. I. Thompson, W. S. Barclay, and M. C. Zambon, "Changes in *in vitro* susceptibility of influenza A H3N2 viruses to a neuraminidase inhibitor drug during evolution in the human host," *J Antimicrob Chemother*, vol. 53, no. 5, pp. 759–65, 2004.

[100] W. P. Glezen and R. B. Couch, "Influenza viruses," in *Viral Infections of Humans: Epidemiology and Control.* (A. F. Evans and R. A. Kaslow, eds.), pp. 473–505, New York: Plenum Medical Book Company, 4th ed., 1997.

[101] C. A. Russell, T. C. Jones, I. G. Barr, N. J. Cox, R. J. Garten, V. Gregory, I. D. Gust, A. W. Hampson, A. J. Hay, A. C. Hurt, J. C. de Jong, A. Kelso, A. I. Klimov, T. Kageyama, N. Komadina, A. S. Lapedes, Y. P. Lin, A. Mosterin, M. Obuchi, T. Odagiri, A. D. M. E. Osterhaus, G. F. Rimmelzwaan, M. W. Shaw, E. Skepner, K. Stohr, M. Tashiro, R. A. M. Fouchier, and D. J. Smith, "The global circulation of seasonal influenza A (H3N2) viruses," *Science*, vol. 320, no. 5874, pp. 340–346, 2008.

[102] J. R. Gog and B. T. Grenfell, "Dynamics and selection of many-strain pathogens," *Proc Natl Acad Sci U S A*, vol. 99, no. 26, pp. 17209–14, 2002.

[103] P. W. Gill and A. M. Murphy, "Naturally acquired immunity to influenza type A: a further prospective study," *Med J Aust*, vol. 2, no. 23, pp. 761–5, 1977.

[104] G. Meiklejohn, T. C. Eickhoff, P. Graves, and J. I, "Antigenic drift and efficacy of influenza virus vaccines, 1976–1977.," *J Infect Dis*, vol. 138, pp. 618–624, Nov 1978.

[105] S. Nakajima, E. Nobusawa, and K. Nakajima, "Variation in response among individuals to antigenic sites on the HA protein of human

influenza virus may be responsible for the emergence of drift strains in the human population," *Virology*, vol. 274, pp. 220–231, Aug 2000.

[106] E. D. Kilbourne, C. Smith, I. Brett, B. A. Pokorny, B. Johansson, and N. Cox, "The total influenza vaccine failure of 1947 revisited: major intrasubtypic antigenic change can explain failure of vaccine in a post-World War II epidemic," *Proc Natl Acad Sci U S A*, vol. 99, no. 16, pp. 10748–52, 2002.

[107] T. A. Reichert, N. Sugaya, D. S. Fedson, W. P. Glezen, L. Simonsen, and M. Tashiro, "The Japanese experience with vaccinating schoolchildren against influenza," *N Engl J Med*, vol. 344, no. 12, pp. 889–96, 2001.

[108] R. Pyhala, R. Visakorpi, N. Ikonen, and M. Kleemola, "Influence of antigenic drift on the intensity of influenza outbreaks: upper respiratory tract infections of military conscripts in Finland," *J Med Virol*, vol. 72, no. 2, pp. 275–80, 2004.

[109] P. D'Agaro, T. Rossi, P. Burgnich, G. Dal Molin, N. Coppola, G. Rocco, and C. Campello, "The molecular epidemiology of influenza viruses: a lesson from a highly epidemic season," *J Clin Pathol*, vol. 61, no. 3, 2008.

[110] F. Ansaldi, G. Icardi, R. Gasparini, C. Campello, S. Puzelli, A. Bella, I. Donatelli, S. Salmaso, and P. Crovari, "New A/H3N2 influenza variant: a small genetic evolution but a heavy burden on the Italian population during the 2004-2005 season," *J Clin Microbiol*, vol. 43, no. 6, pp. 3027–9, 2005.

[111] D. Turner, A. Wailoo, K. Nicholson, N. Cooper, A. Sutton, and K. Abrams, "Appendix 20: Effectiveness of vaccine.," in *Systematic review and economic decision modelling for the prevention and treatment of influenza A and B.*, vol. 7(35), pp. 249–253, Health Technology Assessment, 2003.

[112] D. M. Skowronski, "Influenza VE depends on vaccine component and circulating subtypes: Protection against hospitalization or death due to laboratory-confirmed influenza," in *Eleventh annual conference on vaccine research*, (Baltimore, Maryland), p. s13, 2008.

[113] P. Mangtani, P. Cumberland, C. R. Hodgson, J. A. Roberts, F. T. Cutts, and A. J. Hall, "A cohort study of the effectiveness of influenza vaccine in older people, performed using the United Kingdom General Practice Research Database," *J Infect Dis*, vol. 190, no. 1, pp. 1–10, 2004.

[114] S. E. Ohmit and A. S. Monto, "Influenza vaccine effectiveness in preventing hospitalization among the elderly during influenza type A and type B seasons," *Int J Epidemiol*, vol. 24, pp. 1240–1248, Dec 1995.

[115] A. S. Monto, F. M. Davenport, J. A. Napier, and T. J. Francis, "Modification of an outbreak of influenza in Tecumseh, Michigan by vaccination of schoolchildren," *J Infect Dis*, vol. 122, no. 1, pp. 16–25, 1970.

[116] A. C. Hayward, R. Harling, S. Wetten, A. M. Johnson, S. Munro, J. Smedley, S. Murad, and J. M. Watson, "Effectiveness of an influenza vaccine programme for care home staff to prevent death, morbidity, and health service use among residents: cluster randomised controlled trial," *BMJ*, vol. 333, no. 7581, p. 1241, 2006.

[117] Y. Hirota, "Ecological fallacy and scepticism about influenza vaccine efficacy in Japan: The Maebashi Study," *Vaccine*, vol. 26, no. 50, pp. 6473–6476, 2008.

[118] "Influenza," in *Immunisation against infectious disease: The Green Book*. (D. Salisbury, M. Ramsay, and K. Noakes, eds.), London: The Stationery Office, 3rd ed., 2006.

[119] S. Butt, N. Zhang, and C. A. Joseph, "Vaccination uptake among the 65 years and over and under 65 years at risk in England 2006-07.," tech. rep., Health Protection Agency Centre for Infections, 2007.

[120] C. Joseph, N. Goddard, and D. Gelb, "Influenza vaccine uptake and distribution in England and Wales using data from the General Practice Research Database, 1989/90-2003/04," *J Public Health (Oxf)*, vol. 27, no. 4, pp. 371–7, 2005.

[121] F. Dijkstra, G. A. Donker, B. Wilbrink, A. B. Van Gageldonk-Lafeber, and M. A. Van Der Sande, "Long time trends in influenza-like illness

and associated determinants in The Netherlands," *Epidemiol Infect*, vol. 137, pp. 473–9, 2009.

[122] D. M. Fleming and A. J. Elliot, "Lessons from 40 years' surveillance of influenza in England and Wales," *Epidemiol Infect*, vol. 136, no. 07, pp. 866–875, 2008.

[123] Y. Jin, K. C. Carriere, G. Predy, D. H. Johnson, and T. J. Marrie, "The association between influenza immunization coverage rates and hospitalization for community-acquired pneumonia in Alberta," *Can J Public Health*, vol. 94, no. 5, pp. 341–345, 2003.

[124] A. K. Jha, S. M. Wright, and J. B. Perlin, "Performance measures, vaccinations, and pneumonia rates among high-risk patients in Veterans Administration health care," *Am J Public Health*, vol. 97, no. 12, pp. 2167–2172, 2007.

[125] Y. C. Li, E. C. Norton, and W. H. Dow, "Influenza and pneumococcal vaccination demand responses to changes in infectious disease mortality," *Health Serv Res*, vol. 39, no. 4 Pt 1, pp. 905–25, 2004.

[126] M. F. Parry, B. Grant, A. Iton, P. D. Parry, and D. Baranowsky, "Influenza vaccination: a collaborative effort to improve the health of the community," *Infect Control Hosp Epidemiol*, vol. 25, no. 11, pp. 929–32, 2004.

[127] C. L. Schumann, N. J. Hoxie, and J. M. Vergeront, "Wisconsin trends in pneumonia and influenza mortality, 1980-2003," *WMJ*, vol. 105, no. 1, pp. 40–6, 2006.

[128] P. M. Francisco, M. R. Donalisio, and R. Lattorre Mdo, "[Impact of influenza vaccination on mortality by respiratory diseases among Brazilian elderly persons]," *Rev Saude Publica*, vol. 39, no. 1, pp. 75–81, 2005.

[129] B. Nunes, I. Falco, A. Machado, E. Rodrigues, and J. M. Falco, "Influenza vaccine coverage and the attack rate of influenza-like illness among the elderly in Portugal: is there a correlation?," *Euro Surveill*, vol. 12, p. E070517.2, May 2007.

[130] Y.-H. Chen, S.-H. Liou, C.-C. Chou, W.-L. Su, C.-H. Loh, and S.-H. Lin, "Influenza and pneumococcal vaccination of the elderly in Taiwan," *Vaccine*, vol. 22, pp. 2806–2811, 2004.

[131] J. L. F. Antunes, E. A. Waldman, C. Borrell, and T. M. Paiva, "Effectiveness of influenza vaccination and its impact on health inequalities," *Int J Epidemiol.*, vol. 36, no. 6, pp. 1319–1326, 2007.

[132] P. Honkanen, E. Laara, R. Pyhala, S. L. Kivela, and P. Helena Makela, "Comparison of two vaccination programmes in preventing influenza-related hospitalization among the elderly during two consecutive seasons," *Scand J Infect Dis*, vol. 38, no. 6-7, pp. 506–11, 2006.

[133] C. Viboud and M. Miller, "Health benefits of universal influenza vaccination strategy," *PLoS Med*, vol. 5, no. 10, p. e216, 2008.

[134] Birmingham Research Unit, "Weekly Returns Service Annual Report 2005," tech. rep., Royal College of General Practitioners., 2005.

[135] D. Fleming, "personal communication.".

[136] D. M. Fleming and A. J. Elliot, "The impact of influenza on the health and health care utilisation of elderly people," *Vaccine*, vol. 23, no. Supplement 1, pp. S1–S9, 2005.

[137] D. M. Fleming and K. W. Cross, "Respiratory syncytial virus or influenza?," *Lancet*, vol. 342, no. 8886-8887, pp. 1507–10, 1993.

[138] D. M. Fleming and D. L. Crombie, "The incidence of common infectious diseases: the weekly returns service of the Royal College of General Practitioners," *Health Trends*, vol. 17, no. 1, pp. 13–6, 1985.

[139] Birmingham Research Unit, "Weekly Returns Service Annual Report 2006," tech. rep., Royal College of General Practitioners., 2006.

[140] S. E. Harcourt, D. E. Edwards, D. M. Fleming, R. L. Smith, and G. E. Smith, "How representative is the population covered by the RCGP spotter practice scheme? Using Geographical Information Systems to assess," *J Public Health (Oxf)*, vol. 26, no. 1, pp. 88–94, 2004.

[141] D. M. Fleming and J. Miles, "The representativeness of sentinel practice networks.," *J Public Health (Oxf)*, Sep 2009.

[142] "Morbidity statistics from general practice. series MB5 No. 3," tech. rep., Office of Population Censuses and Surveys, London: Her Majesty's Stationery Office, 1995.

[143] L. Rickards, K. Fox, C. Roberts, L. Fletcher, and E. Goddard, "Living in Britain: Results from the 2002 General Household Survey," tech. rep., The Stationery Office., 2002.

[144] Birmingham Research Unit, "The Weekly Returns Service: Annual Report 1999," tech. rep., Royal College of General Practitioners., 1999.

[145] D. M. Fleming, K. W. Cross, D. L. Crombie, and R. J. Lancashire, "Respiratory illness and mortality in England and Wales. A study of the relationships between weekly data for the incidence of respiratory disease presenting to general practitioners, and registered deaths," *Eur J Epidemiol*, vol. 9, no. 6, pp. 571–6, 1993.

[146] Birmingham Research Unit, "The Weekly Returns Service: Annual Report 2000," tech. rep., Royal College of General Practitioners., 2000.

[147] T. Devis and C. Rooney, "Death certification and the epidemiologist.," *Health Stat Quarterly*, no. 01, pp. 21–33, 1999.

[148] A. Brock, C. Griffiths, and C. Rooney, "The impact of introducing ICD-10 on analysis of respiratory mortality trends in England and Wales," *Health Stat Quarterly*, vol. 29, pp. 9–17, 2006.

[149] C. Viboud, O. N. Bjornstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell, "Synchrony, waves, and spatial hierarchies in the spread of influenza," *Science*, vol. 312, no. 5772, pp. 447–51, 2006.

[150] L. Simonsen, "The global impact of influenza on morbidity and mortality," *Vaccine*, vol. 17 Suppl 1, pp. S3–10, 1999.

[151] T. A. Reichert, L. Simonsen, A. Sharma, S. A. Pardo, D. S. Fedson, and M. A. Miller, "Influenza and the winter increase in mortality in the United States, 1959-1999," *Am J Epidemiol*, vol. 160, no. 5, pp. 492–502, 2004.

[152] J. Housworth and A. D. Langmuir, "Excess mortality from epidemic influenza, 1957-1966," *Am J Epidemiol*, vol. 100, no. 1, pp. 40–48, 1974.

[153] "Stata statistical software: Release 9.." College Station, TX: StataCorp LP., 2005.

[154] C. Rooney, C. Griffiths, and L. Cook, "The implementation of ICD-10 for cause of death coding - some preliminary results from the bridge coding study," *Health Stat Quarterly*, vol. 13, pp. 31–41., 2002.

[155] N. S. Crowcroft, F. Cutts, and M. C. Zambon, "Respiratory syncytial virus: an underestimated cause of respiratory infection, with prospects for a vaccine," *Commun Dis Public Health*, vol. 2, no. 4, pp. 234–41, 1999.

[156] C. P. Farrington, N. J. Andrews, A. D. Beale, and M. A. Cathpole, "A statistical algorithm for the early detection of outbreaks of infectious disease," *J R Stat Soc A*, vol. 159, no. 3, pp. 547–563, 1996.

[157] H. Jick, S. S. Jick, and L. E. Derby, "Validation of information recorded on general practitioner based computerised data resource in the United Kingdom," *BMJ*, vol. 302, no. 6779, pp. 766–8, 1991.

[158] M. S. Pereira and P. Chakraverty, "Influenza in the United Kingdom 1977-1981," *J Hyg (Lond)*, vol. 88, no. 3, pp. 501–12, 1982.

[159] M. Pereira, F. A. Assaad, and P. J. Delon, "Influenza surveillance.," *Bull World Health Organ*, vol. 56, no. 2, pp. 192–203, 1978.

[160] P. Chakraverty, P. Cunningham, G. Z. Shen, and M. S. Pereira, "Influenza in the United Kingdom 1982-85," *J Hyg (Lond)*, vol. 97, no. 2, pp. 347–58, 1986.

[161] D. J. Smith, S. Forrest, D. H. Ackley, and A. S. Perelson, "Variable efficacy of repeated annual influenza vaccination," *Proc Natl Acad Sci U S A*, vol. 96, no. 24, pp. 14001–6, 1999.

[162] "Recommended composition of influenza virus vaccines for use in the 1989-1990 season," *Wkly Epidemiol Rec*, vol. 64, no. 8, pp. 53–60, 1989.

[163] "Recommended composition of influenza virus vaccines for use in the 1990-1991 season," *Wkly Epidemiol Rec*, vol. 65, no. 8, pp. 53–6, 1990.

[164] C. A. Joseph, D. Dedman, K. Fern, P. Chakraverty, and J. M. Watson, "Influenza surveillance in England and Wales: November 1991-June 1992," *Commun Dis Rep CDR Rev*, vol. 2, no. 13, pp. R149–52, 1992.

[165] D. Dedman, C. A. Joseph, P. Chakraverty, and J. M. Watson, "Influenza surveillance, England and Wales: October 1992-June 1993," *Commun Dis Rep CDR Rev*, vol. 3, no. 13, pp. R184–6, 1993.

[166] D. J. Dedman, C. A. Joseph, P. Chakraverty, D. M. Fleming, and J. M. Watson, "Influenza surveillance, England and Wales: October 1993 to June 1994," *Commun Dis Rep CDR Rev*, vol. 4, no. 13, pp. R164–8, 1994.

[167] E. J. Hutchinson, C. A. Joseph, P. Chakraverty, M. Zambon, D. M. Fleming, and J. M. Watson, "Influenza surveillance in England and Wales: October 1994 to June 1995," *Commun Dis Rep CDR Rev*, vol. 5, no. 13, pp. R200–4, 1995.

[168] E. J. Hutchinson, C. A. Joseph, M. Zambon, D. M. Fleming, and J. M. Watson, "Influenza surveillance in England and Wales: October 1995 to June 1996," *Commun Dis Rep CDR Rev*, vol. 6, no. 12, pp. R163–9, 1996.

[169] D. J. Dedman, C. A. Joseph, M. Zambon, D. M. Fleming, and J. M. Watson, "Influenza surveillance in England and Wales: October 1996 to June 1997," *Commun Dis Rep CDR Rev*, vol. 7, no. 13, pp. R212–9, 1997.

[170] D. J. Dedman, M. Zambon, P. V. Buynder, D. M. Fleming, J. M. Watson, and C. A. Joseph, "Influenza surveillance in England and Wales: October 1997 to June 1998," *Commun Dis Public Health*, vol. 1, no. 4, pp. 244–51, 1998.

[171] P. Whiting, C. A. Joseph, M. Zambon, M. Nunn, D. Fleming, and J. M. Watson, "Influenza activity in England and Wales: October 1998 to June 1999," *Commun Dis Public Health*, vol. 2, no. 4, pp. 273–9, 1999.

[172] N. L. Goddard, C. A. Joseph, M. Zambon, M. Nunn, D. Fleming, and J. M. Watson, "Influenza surveillance in England and Wales: October 1999 to May 2000," *Commun Dis Public Health*, vol. 3, no. 4, pp. 261–6, 2000.

[173] N. L. Goddard, C. A. Joseph, M. Zambon, M. Nunn, D. Fleming, and J. M. Watson, "Influenza surveillance in the United Kingdom: October 2000 to May 2001," *Commun Dis Rep CDR Suppl*, pp. 1–7, 2001.

[174] J. P. Crofts, N. L. Goddard, C. A. Joseph, M. Zambon, J. Ellis, D. M. Fleming, and J. M. Watson, "Influenza surveillance in the United Kingdom: October 2001 to May 2002," *Commun Dis Rep CDR Suppl*, pp. 1–7, 2002.

[175] J. P. Crofts, C. A. Joseph, M. Zambon, J. Ellis, D. M. Fleming, and J. M. Watson, "Influenza survellance in the United Kingdom: October 2002 to May 2003," *Commun Dis Rep CDR Suppl*, vol. 14, no. 10, pp. 1–9, 2004.

[176] M. K. Cooke, J. P. Crofts, C. A. Joseph, N. L. Goddard, J. Ellis, M. Zambon, D. M. Fleming, J. Nguyen-Van-Tam, and J. M. Watson, "Influenza and other respiratory viruses surveillance in the United Kingdom: October 2003 to May 2004," *Commun Dis Rep CDR Suppl*, vol. 15, no. 3, pp. 1–8, 2005.

[177] H. Zhao, M. K. Cooke, C. A. Joseph, J. Ellis, M. Zambon, D. M. Fleming, J. S. Nguyen-van Tam, and J. M. Watson, "Surveillance of influenza and other respiratory viruses in the United Kingdom: October 2004 to May 2005," *Commun Dis Rep CDR Suppl*, vol. 16, no. 27, pp. 1–8, 2006.

[178] A. Rambaut, O. G. Pybus, M. I. Nelson, C. Viboud, J. K. Taubenberger, and E. C. Holmes, "The genomic and epidemiological dynamics of human influenza A virus," *Nature*, vol. 453, no. 7195, pp. 615–9, 2008.

[179] Standing Medical Advisory Committee Subcommittee on Antibiotic Resistance, "The Path of Least Resistance," tech. rep., Department of Health, 1998.

[180] A. J. Dobson, *An introduction to generalized linear models.* Texts in Statistical Science, Boca Raton: Chapman & Hall/CRC, 2nd ed., 2002.

[181] W. Venables and B. Ripley, *Modern Applied Statistics with S.* Statistics and Computing, New York, New York: Springer Science+Business Media, LLC, 4th ed., 2002.

[182] B. Kirkwood and J. Sterne, *Essential medical statistics.* Malden: Blackwell Publishing, 2nd ed., 1988.

[183] Office for National Statistics, "Mid-year population estimates." http://www.statistics.gov.uk/, Accessed 21 November 2007.

[184] D. L. Schanzer, J. M. Langley, and T. W. S. Tam, "Role of influenza and other respiratory viruses in admissions of adults to Canadian hospitals," *Influenza Other Respi Viruses*, vol. 2, no. 1, pp. 1–8, 2008.

[185] S. Hajat, W. Bird, and A. Haines, "Cold weather and GP consultations for respiratory conditions by elderly people in 16 locations in the UK," *Eur J Epidemiol*, vol. 19, no. 10, pp. 959–968, 2004.

[186] T. J. Hastie, "Generalized additive models," in *Statistical Models in S* (J. M. Chambers and T. J. Hastie, eds.), Pacific Grove, California: Wadsworth & Brooks/Cole, 1992.

[187] H. Akaike, "Information theory and an extension of the maximum likelihood principle.," in *Second International Symposium on Information Theory.* (B. N. Petrov and F. Csaki, eds.), (Budapest: Akademiai Kiado.), pp. 267–281, 1973.

[188] I. L. MacDonald and W. Zucchini, *Hidden Markov and Other Models for Discrete-valued Time Series.* London: Chapman and Hall., 1997.

[189] W. Yu and G. B. Sheble, "Modeling electricity markets with hidden Markov models," *Electric Power Systems Research*, vol. 76, no. 6-7, pp. 445–451, 2006.

[190] A. Gelman, "Markov chain Monte Carlo in Practice," ch. Inference and monitoring convergence, pp. 131–143, Chapman & Hall/CRC, 1996.

[191] W. Hastings, "Monte Carlo sampling methods using Marko chains and systems," *J R Stat Soc B*, vol. 56, pp. 549–603, 1994.

[192] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans Pattn Anal Mach Intell*, vol. 6, pp. 721–741, 1984.

[193] D. J. Spiegelhalter, K. Abrams, and J. Myles, *Bayesian Approaches to Clinical Trials and Health Care Evaluation*. Chichester: John Wiley & Sons, Ltd., 2004.

[194] A. O'Hagan, "Science, subjectivity and software (comment on articles by Berger and by Goldstein)," *Bayesian Analysis*, vol. 3, no. 1, pp. 445–50, 2006.

[195] J. E. Oakley and A. O'Hagan, "Uncertainty in prior elicitations: a nonparametric approach," *Biometrika*, vol. 94, no. 2, pp. 427–441, 2007.

[196] P. H. Garthwaite, J. B. Kadane, and A. O'Hagan, "Statistical methods for eliciting probability distributions," *J Am Stat Assoc*, vol. 100, no. 470, pp. 680–701, 2005.

[197] D. F. Stroup, S. B. Thacker, and J. L. Herndon, "Application of multiple time series analysis to the estimation of pneumonia and influenza mortality by age 1962-1983," *Stat Med*, vol. 7, no. 10, pp. 1045–59, 1988.

[198] A. B. Sabin, "Mortality from pneumonia and risk conditions during influenza epidemics. High influenza morbidity during nonepidemic years," *JAMA*, vol. 237, no. 26, pp. 2823–8, 1977.

[199] V. H. Menec, C. Black, L. MacWilliam, and F. Y. Aoki, "The impact of influenza-associated respiratory illnesses on hospitalizations, physician visits, emergency room visits, and mortality," *Can J Public Health*, vol. 94, no. 1, pp. 59–63, 2003.

[200] S. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations," *J Comput Graph Stat*, vol. 7, no. 4, pp. 434–455, 1998.

[201] A. Gelman, X. L. Meng, and H. Stern, "Posterior predictive assessment of model fitness via realized discrepancies.," *Statistica Sinica*, vol. 6, pp. 703–807, 1996.

[202] E. C. Marshall and D. J. Spiegelhalter, "Identifying outliers in Bayesian hierarchical models: a simulation-based approach.," *Bayesian Analysis*, vol. 2, no. 2, pp. 409–44, 2007.

[203] A. S. Mugglin, N. Cressie, and I. Gemmell, "Hierarchical statistical modelling of influenza epidemic dynamics in space and time," *Stat Med*, vol. 21, no. 18, pp. 2703–21, 2002.

[204] C. Chatfield, *The analysis of time series: an introduction.* Texts in Statistical Science, Boca Raton: Chapman & Hall/CRC, 6th ed., 2004.

[205] P. Congdon, *Bayesian Statistical Modelling.* Wiley Series in Probability and Statistics, West Sussex: John Wiley & Sons, Ltd., 2nd ed., 2006.

[206] T. C. Lin and M. Pourahmadi, "Nonparametric and non-linear models and data mining in time series: a case-study on the Canadian lynx data," *J R Stat Soc C*, vol. 47, no. 2, pp. 187–201, 1998.

[207] A. Thomas, B. O Hara, U. Ligges, and S. Sturtz, "Making BUGS Open," *R News*, vol. 6, no. 1, pp. 12–17, 2006.

[208] B. Lopman, "personal communication.".

[209] "Influenza vaccine uptake results by age and PCT." http://www.immunisation.nhs.uk/Vaccines/Flu/Resources. Accessed 1 November 2007.

[210] R. S. Chapman, G. E. Smith, F. Warburton, R. T. Mayon-White, and D. M. Fleming, "Impact of NHS Direct on general practice consultations during the winter of 1999-2000: analysis of routinely collected data," *BMJ*, vol. 325, no. 7377, pp. 1397–8, 2002.

[211] K. Dunnell, "Ageing and mortality in the UK: National Statistician's annual article on the population," *Popul Trends*, vol. 134, pp. 6–23, 2008.

[212] C. L. Trotter, J. M. Stuart, R. George, and E. Miller, "Increasing hospital admissions for pneumonia, England," *Emerg Infect Dis*, vol. 14, no. 5, pp. 727–33, 2008.

[213] J. R. Panickar, S. R. Dodd, R. L. Smyth, and J. M. Couriel, "Trends in deaths from respiratory illness in children in England and Wales from 1968 to 2000," *Thorax*, vol. 60, no. 12, pp. 1035–8, 2005.

[214] A. Ozonoff, "personal communication.".

[215] N. Alexander, J. Bethony, R. Corra-Oliveira, L. C. Rodrigues, P. Hotez, and S. Brooker, "Repeatability of paired counts.," *Stat Med*, vol. 26, pp. 3566–3577, Aug 2007.

[216] D. A. Foster, A. Talsma, A. Furumoto-Dawson, S. E. Ohmit, J. R. Margulies, N. H. Arden, and A. S. Monto, "Influenza vaccine effectiveness in preventing hospitalization for pneumonia in the elderly.," *Am J Epidemiol*, vol. 136, pp. 296–307, Aug 1992.

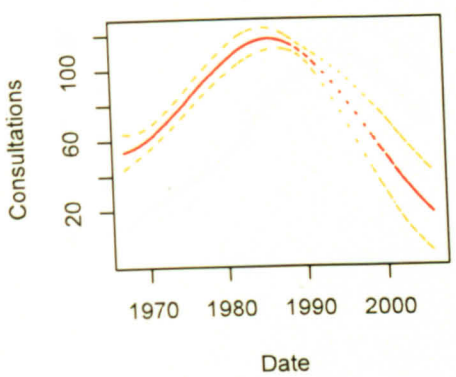
[217] C. P. Farrington, "Estimation of vaccine effectiveness using the screening method.," *Int J Epidemiol*, vol. 22, pp. 742–746, Aug 1993.

[218] D. M. Fleming, N. J. Andrews, J. S. Elllis, A. Bermingham, P. Sebastianpillai, A. J. Elliot, E. Miller, and M. Zambon, "Estimating influenza vaccine effectiveness using routinely collected laboratory data," *J Epidemiol Community Health*, 2009.

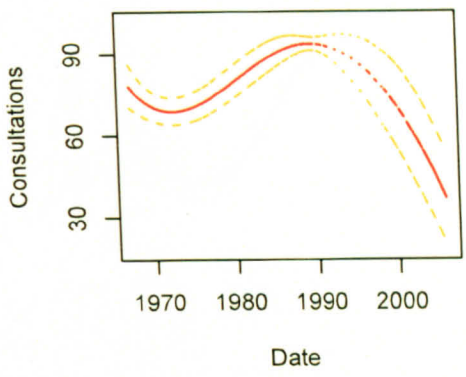
Appendices

A. Trends in other GP consultations

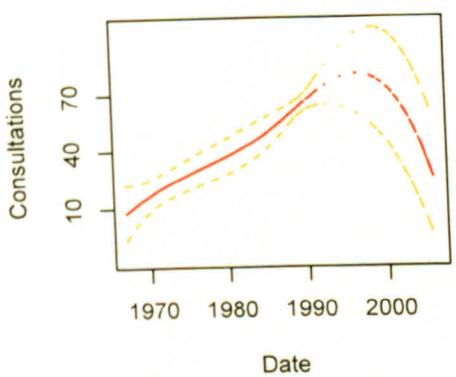
Similar declines to those in ILI since 2000 were observed for other upper and lower respiratory tract infections, as well as non-respiratory consultation categories (figures for upper respiratory tract infections: A.1 to A.4, figures for lower respiratory tract infections: A.5 to A.7, figure for non-respiratory infections: A.8).



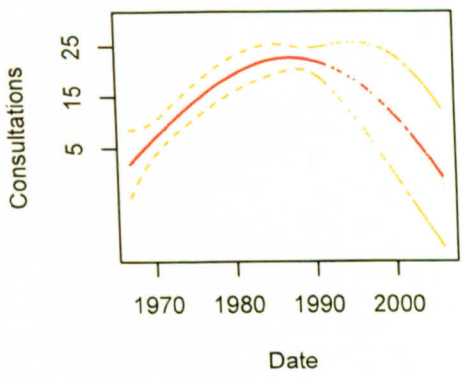
(a) 0-4 years



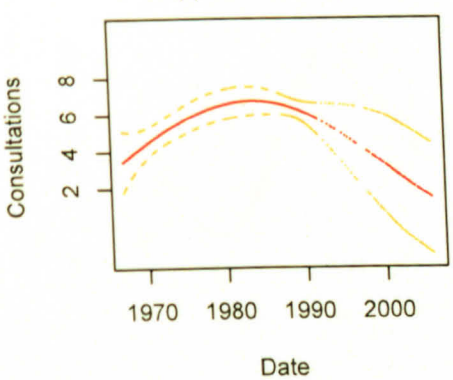
(b) 5-14



(c) 15-44

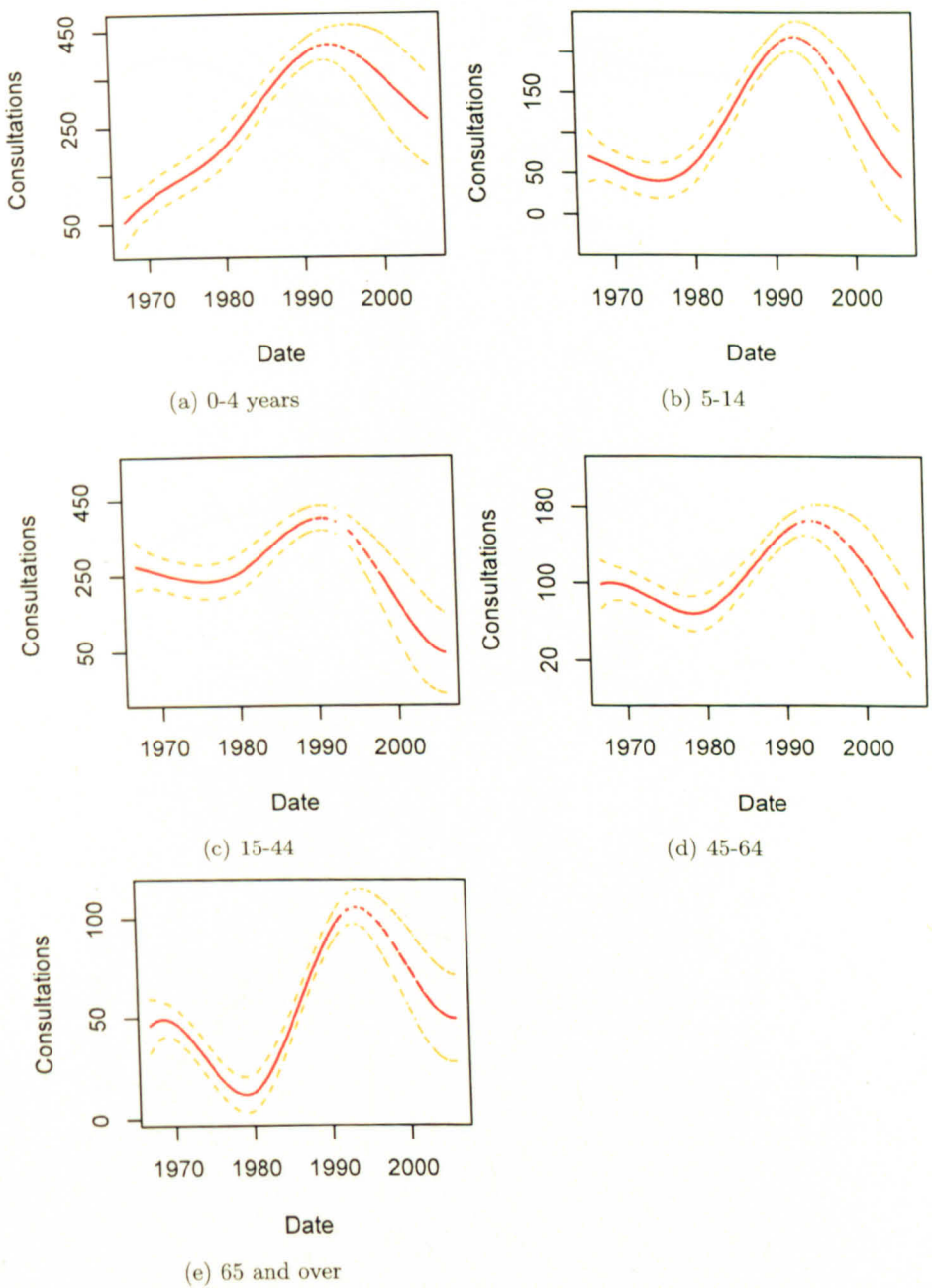


(d) 45-64



(e) 65 and over

Figure A.1.: **Long-term trend in GP consultations for otitis media in those aged (a) 0-4 years, (b) 5-14 years, (c) 15-44 years, (d) 45-64 years and (e) 65 years of age and over.** 25% of the highest counts were excluded from model fitting.



(a) 0-4 years

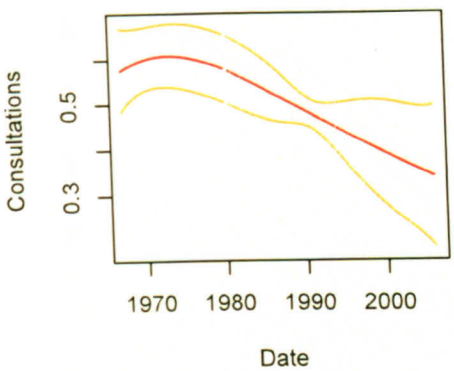
(b) 5-14

(c) 15-44

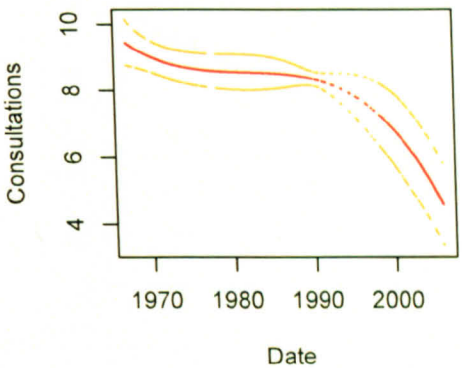
(d) 45-64

(e) 65 and over

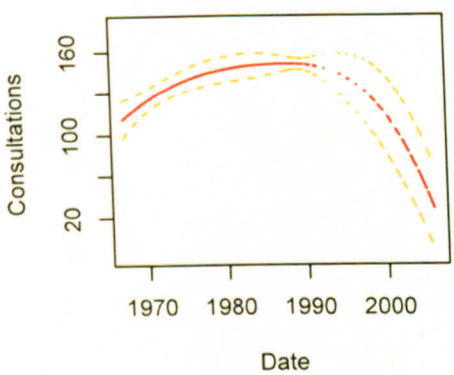
Figure A.2.: **Long-term trend in GP consultations for common cold in those aged (a) 0-4 years, (b) 5-14 years, (c) 15-44 years, (d) 45-64 years and (e) 65 years of age and over.** 15% of the highest counts were excluded from model fitting.



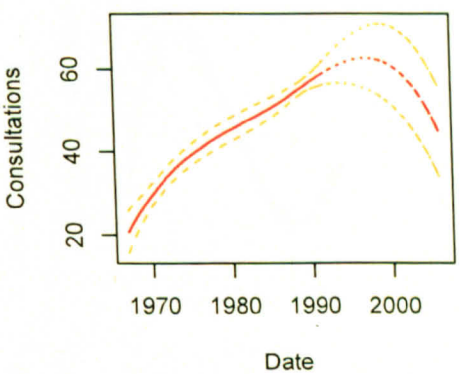
(a) 0-4 years



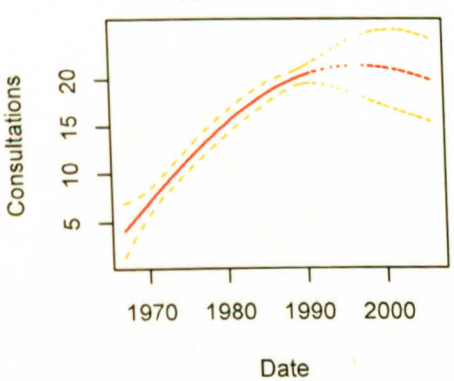
(b) 5-14



(c) 15-44

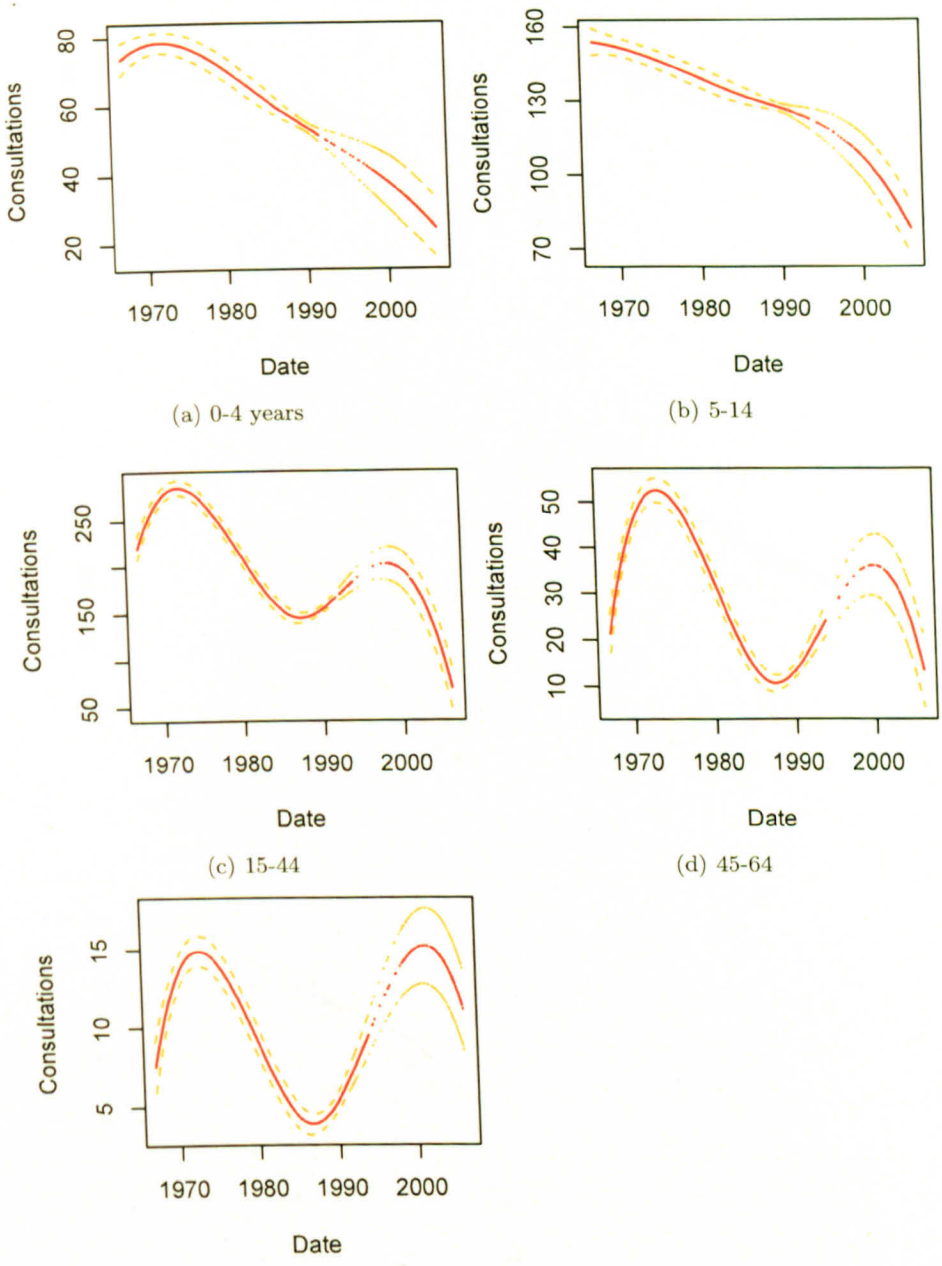


(d) 45-64



(e) 65 and over

Figure A.3.: **Long-term trend in GP consultations for acute sinusitis in those aged (a) 0-4 years, (b) 5-14 years, (c) 15-44 years, (d) 45-64 years and (e) 65 years of age and over.** 25% of the highest counts were excluded from model fitting.



(a) 0-4 years

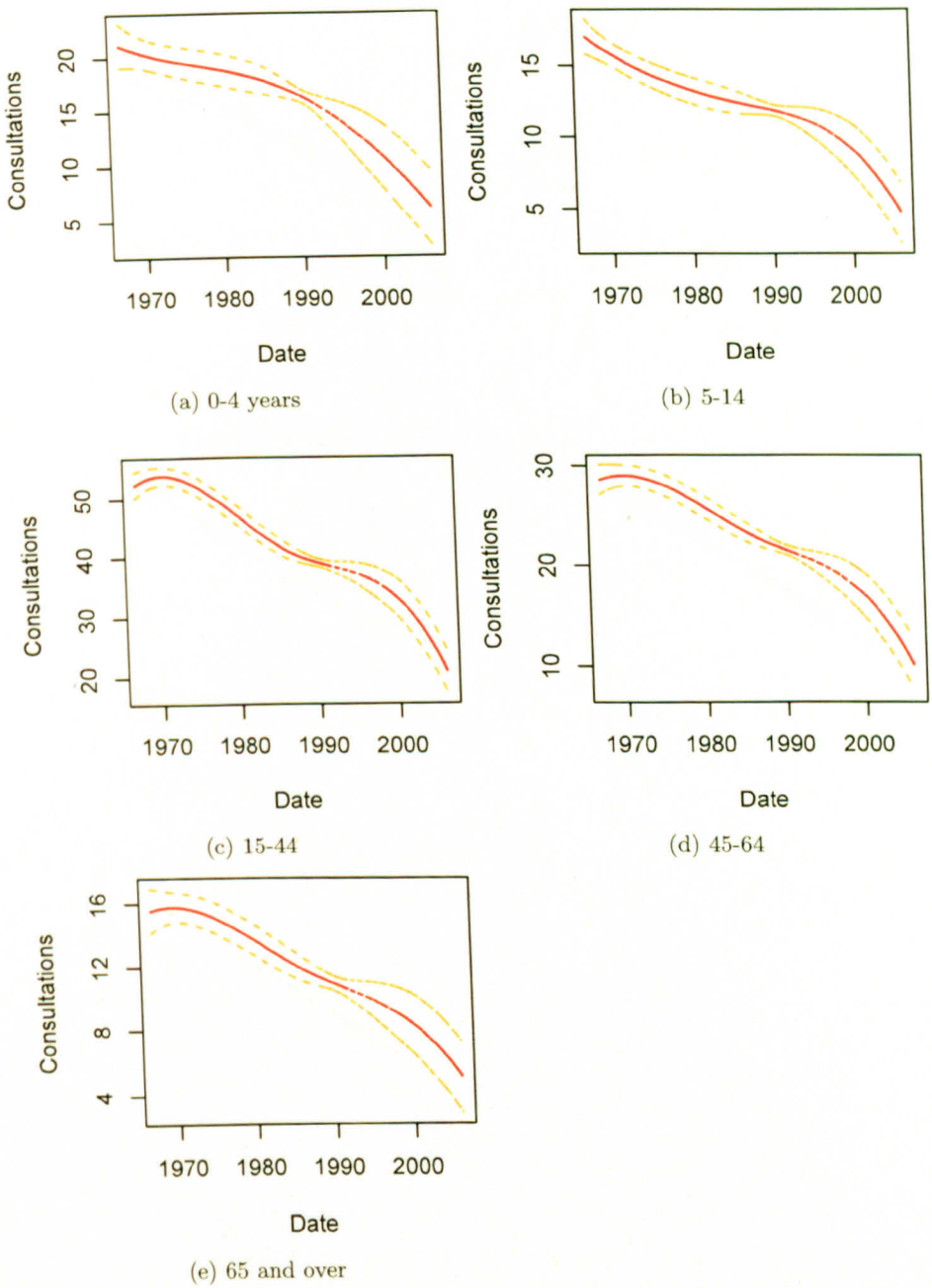
(b) 5-14

(c) 15-44

(d) 45-64

(e) 65 and over

Figure A.4.: **Long-term trend in GP consultations for acute tonsillitis in those aged (a) 0-4 years, (b) 5-14 years, (c) 15-44 years, (d) 45-64 years and (e) 65 years of age and over.** 15% of the highest counts were excluded from model fitting.



(a) 0-4 years

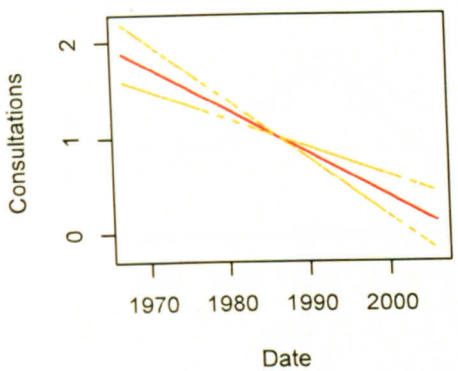
(b) 5-14

(c) 15-44

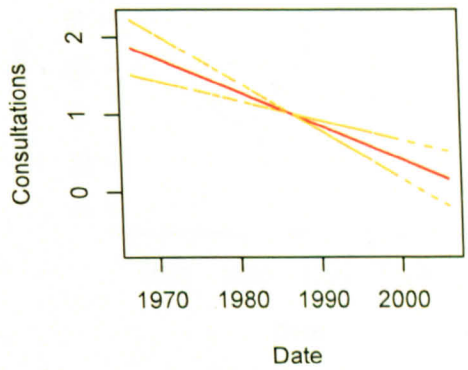
(d) 45-64

(e) 65 and over

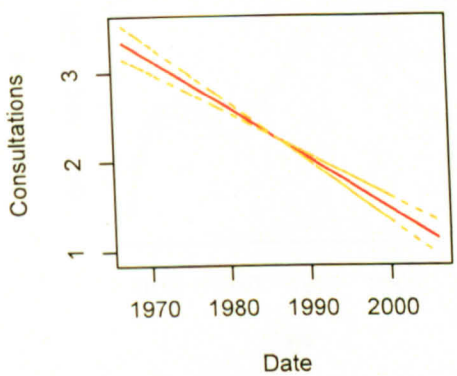
Figure A.5.: **Long-term trend in GP consultations for laryngitis in those aged (a) 0-4 years, (b) 5-14 years, (c) 15-44 years, (d) 45-64 years and (e) 65 years of age and over.** 10% of the highest counts were excluded from model fitting.



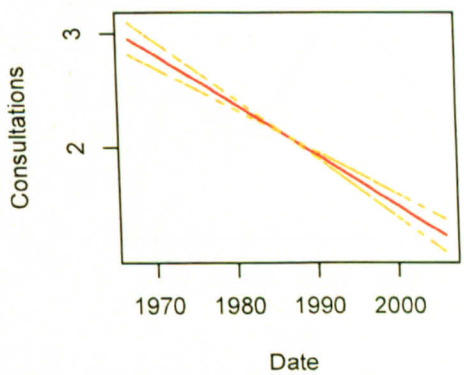
(a) 0-4 years



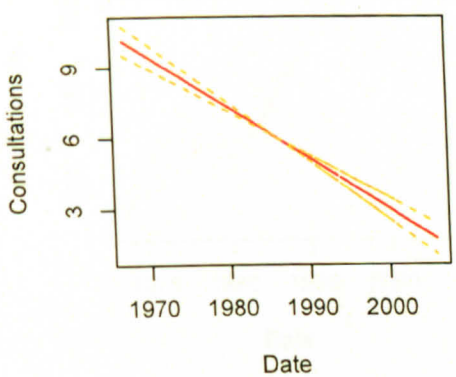
(b) 5-14



(c) 15-44

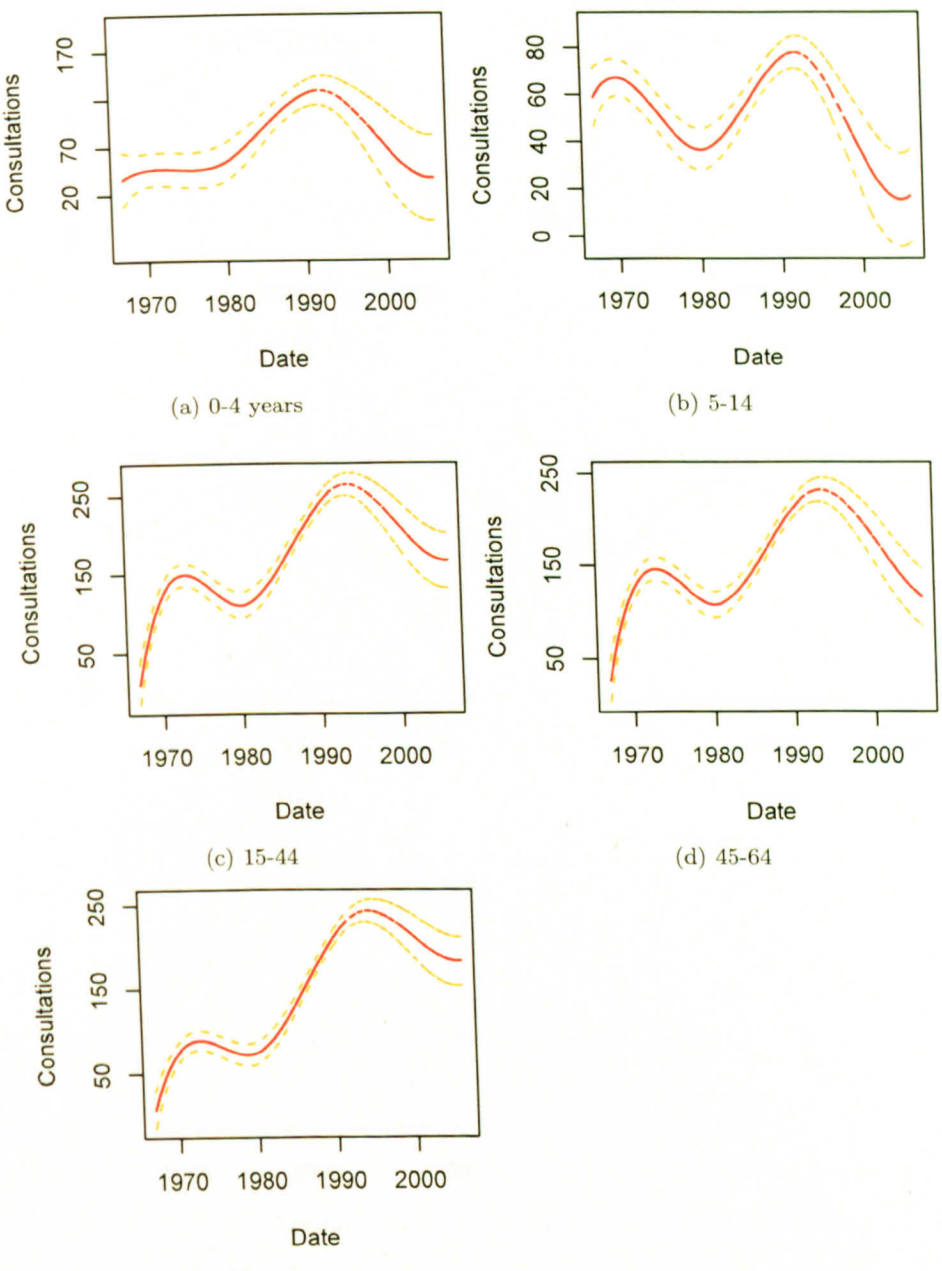


(d) 45-64



(e) 65 and over

Figure A.6.: **Long-term trend in GP consultations for pneumonia in those aged (a) 0-4 years, (b) 5-14 years, (c) 15-44 years, (d) 45-64 years and (e) 65 years of age and over.** 5% of the highest counts were excluded from model fitting.



(a) 0-4 years

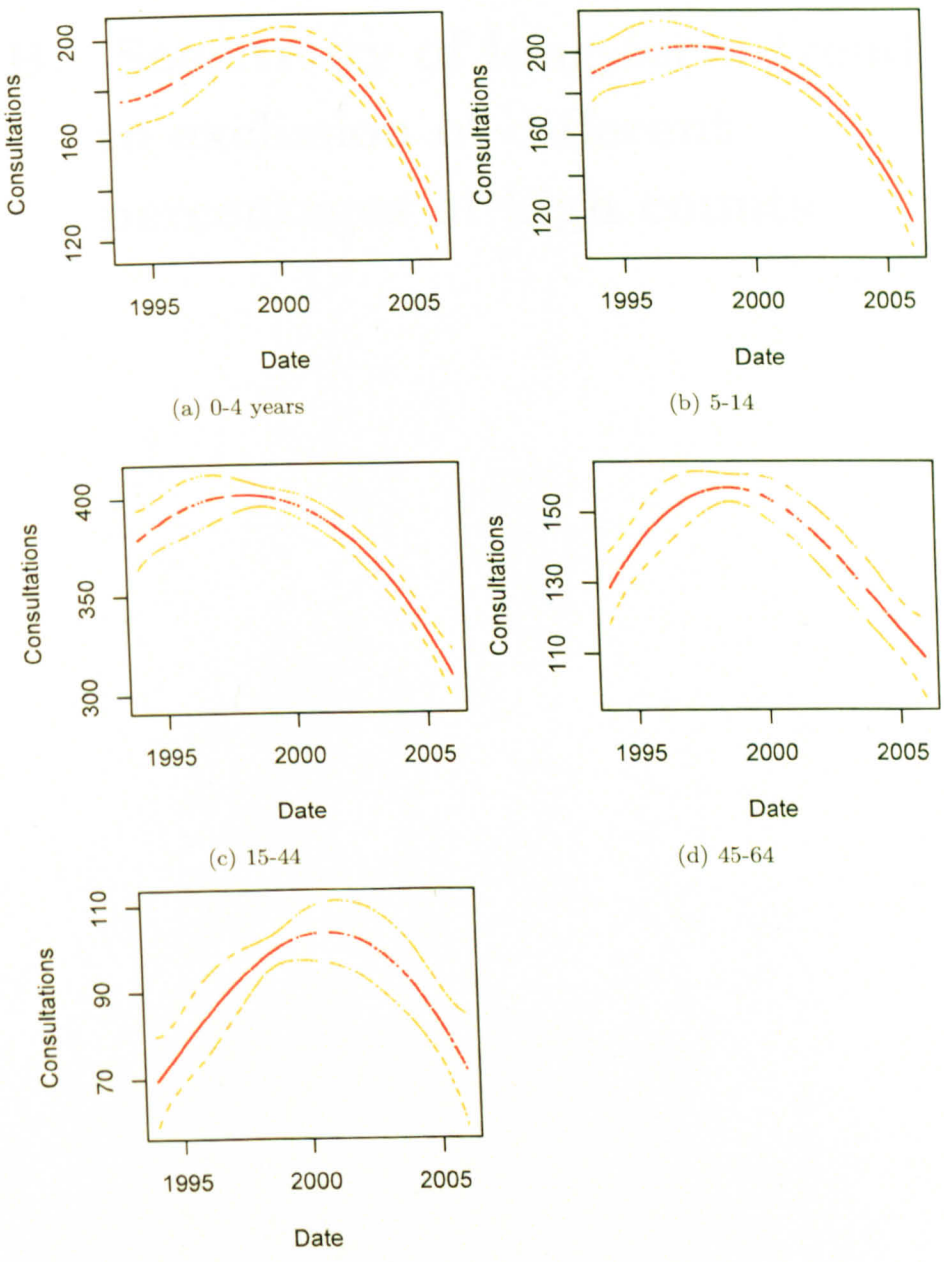
(b) 5-14

(c) 15-44

(d) 45-64

(e) 65 and over

Figure A.7.: **Long-term trend in GP consultations for acute bronchitis in those aged (a) 0-4 years, (b) 5-14 years, (c) 15-44 years, (d) 45-64 years and (e) 65 years of age and over.** 10% of the highest counts were excluded from model fitting.



(a) 0-4 years

(b) 5-14

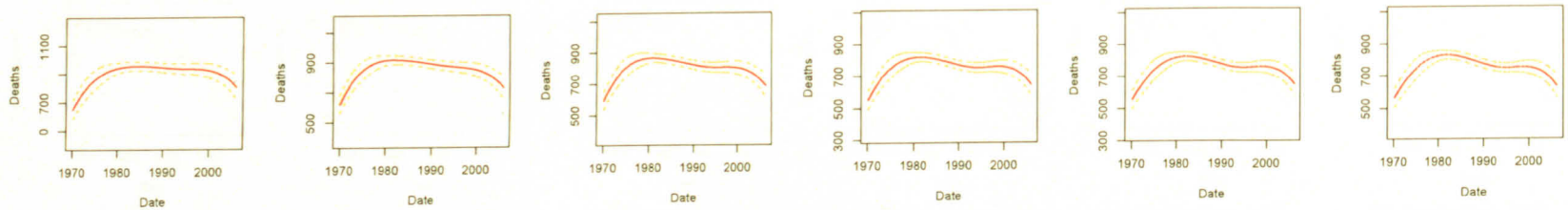
(c) 15-44

(d) 45-64

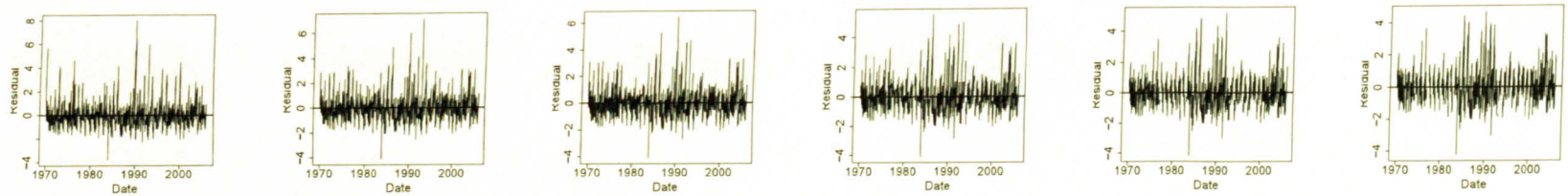
(e) 65 and over

Figure A.8.: **Long-term trend in GP consultations for infectious and parasitic diseases, which include most infectious disease consultations apart from respiratory infections, in those aged (a) 0-4 years, (b) 5-14 years, (c) 15-44 years, (d) 45-64 years and (e) 65 years of age and over.** 25% of the highest counts were excluded from model fitting.

B. Sensitivity of long-term trend to exclusion of different percentages of high counts

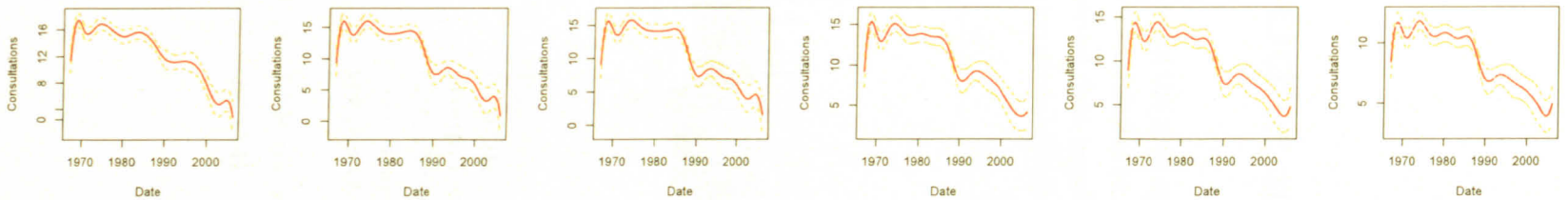


(a) Excluding 1% (b) Excluding 2.5% (c) Excluding 5% (d) Excluding 10% (e) Excluding 15% (f) Excluding 20%

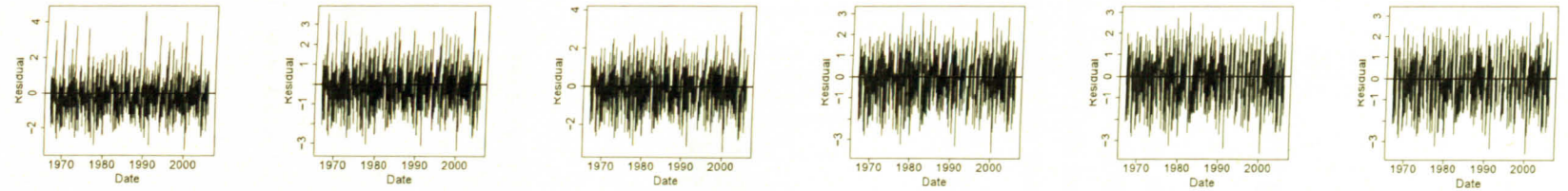


(g) Excluding 1% (h) Excluding 2.5% (i) Excluding 5% (j) Excluding 10% (k) Excluding 15% (l) Excluding 20%

Figure B.1.: **Top panel: spline component of fitted long-term trend in P&I in those $\geq$ 65 (solid curve) and a 95% CI about the fitted spline (dashed curves either side of the fitted curve) excluding (a) 1, (b) 2.5, (c) 5, (d) 10, (e) 15 and (f) 20% of the highest counts. Bottom panel: residuals (observed minus fitted value each week) from these same models((g) - (l)).**



(a) Excluding 1% (b) Excluding 2.5% (c) Excluding 5% (d) Excluding 10% (e) Excluding 15% (f) Excluding 20%

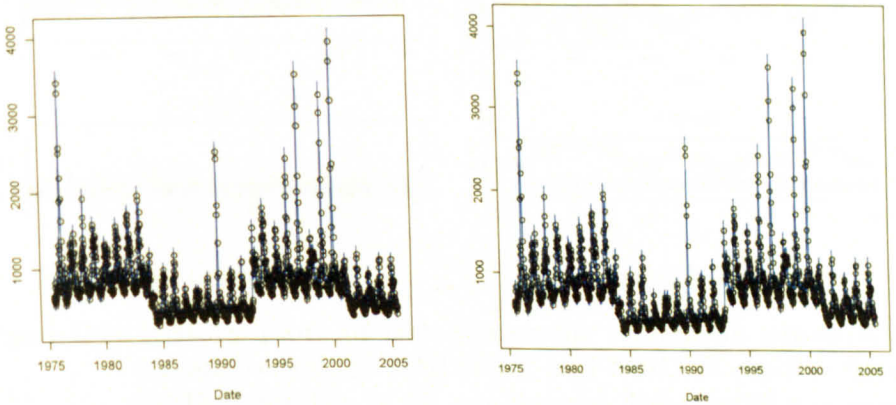


(g) Excluding 1% (h) Excluding 2.5% (i) Excluding 5% (j) Excluding 10% (k) Excluding 15% (l) Excluding 20%

Figure B.2.: **The top row of plots are the long-term trend in ILI in those 65 years of age and older (solid curve) and a 95% CI about the fitted spline (dashed curves either side of the fitted curve) excluding (a) 1, (b) 2.5, (c) 5, (d) 10, (e) 15 and (f) 20% of the highest counts from model fitting. The bottom row are plots of the model residuals (observed minus fitted value each week) from these same models((g) - (l))**

C. Other model structures

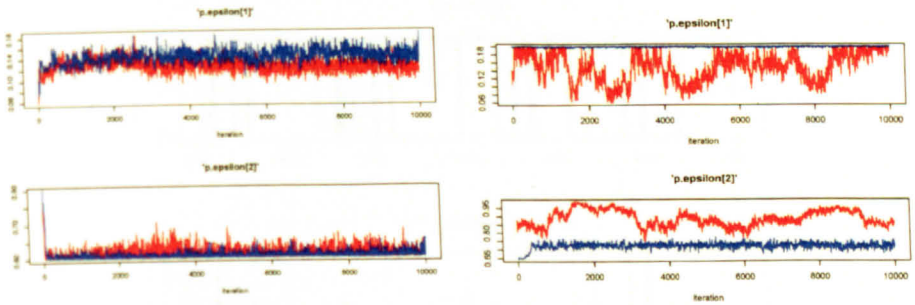
P&I data for the $\geq$ 65 age group and ILI data for the 15-44 age group are underdispersed relative to the NBLOGR and NBIDR models (e.g. figure C.1). No observed data points fall outside 95% CrIs for model-predicted counts. Transition probabilities from these models do not converge (figure C.2).



(a) NBLOGR (b) NBIDR

Figure C.1.: **Posterior predictive density plots of NBLOGR and NBIDR models fitted to P&I from the $\geq$ 65 age group.** The predicted counts and a CrI for each predicted count (lines) are plotted on the same graph as the observed data (circles). These data are underdispersed (no observed data outside the posterior predicted CrIs) relative to both models.

SQR and LNR models fit P&I data from the $\geq$ 65 age group poorly (figures C.3 and C.4).



(a) NBLOGR (b) NBIDR

Figure C.2.: **History plots of two chains for transition probability parameters from NBLOGR and NBIDR models fitted to P&I for the $\geq$ 65 age group.** For identity-link model (NBIDR) chains do not mix well - 1 chain moves widely around the parameter space and does not appear to settle down to a value; the two chains do not appear to be moving towards a common parameter space. Two chains for transition probabilities from NBLOGR model mix better.

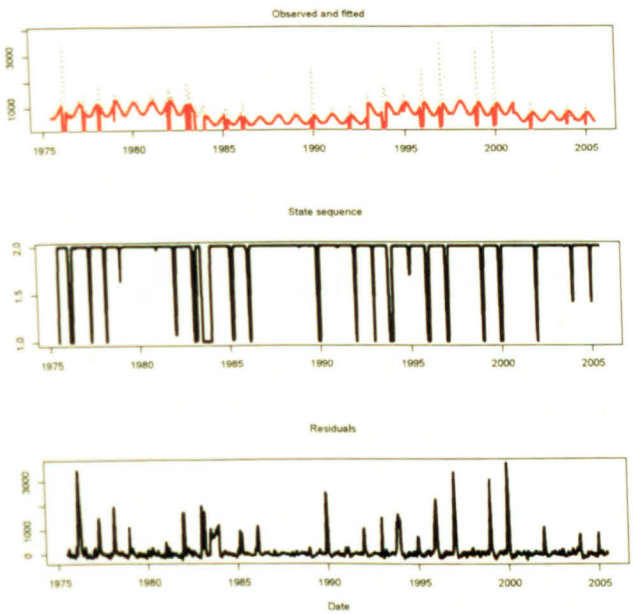


Figure C.3.: **State-sequence for SQR model fitted to P&I.** Top panel: observed (dashed) and fitted P&I data for the ≥ 65 age group (solid); middle panel: state sequence (1 is the 'normal' state, 2 the 'aberrant' state); bottom panel: residuals (observed minus fitted P&I count for each week). The state sequence shown was plotted by averaging the state sequence estimated by each of the two chains. The model-predicted time series is very different from the observed time series. Residual plot shows that model-predicted time series captured little of the variability in the data. Labels for state variables have also switched whereby state label 1 now refers to the 'aberrant' state and label 2 to the 'normal' state; this is indicative of a poor model.

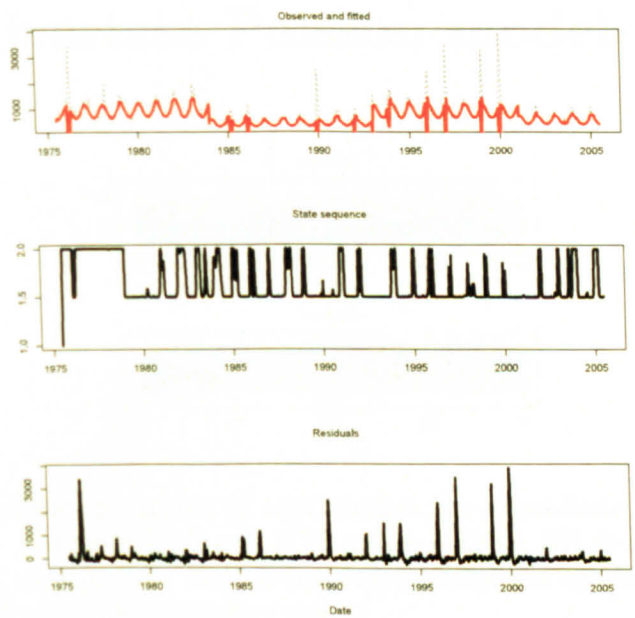


Figure C.4.: **State-sequence for LNR models fitted to P&I for the ≥ 65 age group.** As with SQR model fit, model-predicted time series is very different from the observed time series and residual plots show that model-predicted time series captured little of the variability in the data. The state sequence is poorly estimated - for most weeks the two chains for the state sequence are not in agreement as to whether a given week is from the 'normal' or 'aberrant' state - because transition probability parameters have not converged. This is evidenced by the state sequence having a value of approximately 0.5 for most weeks; this is an average state sequence across two chains which do not agree on state assignment for most weeks.

D. Fit and convergence of univariate Poisson models

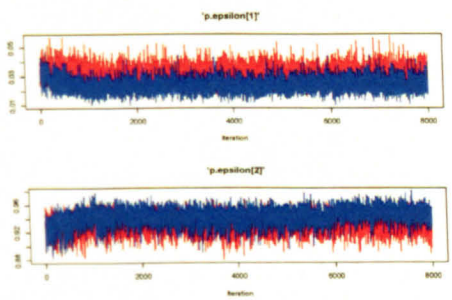


Figure D.1.: **History plots of two chains for transition probability parameters from IDR model fit to ILI from the $\geq$ 65 age group.** Both transition probabilities appear to converge to similar parameter space.

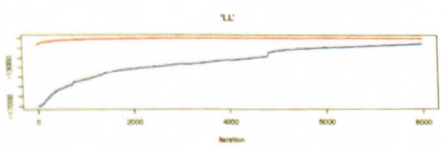
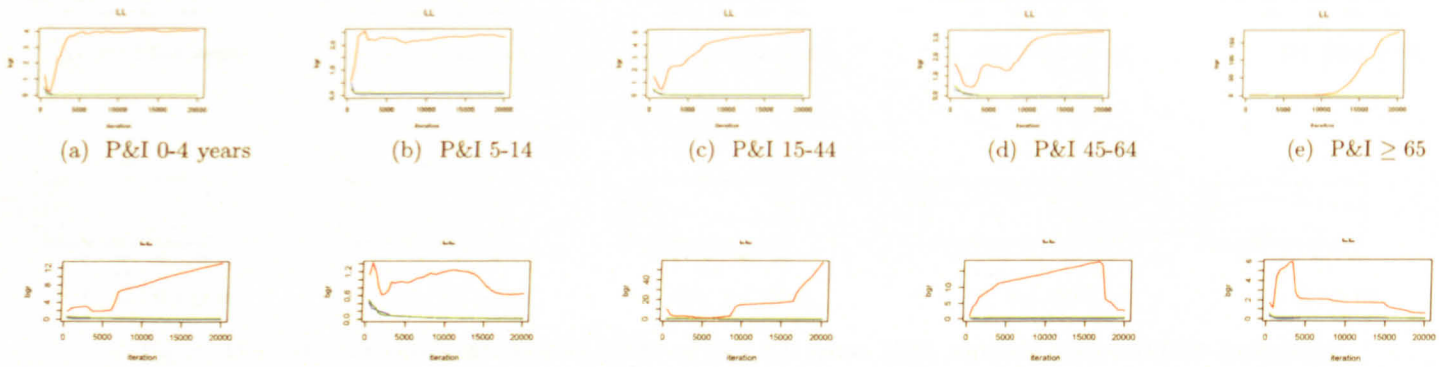


Figure D.2.: **History plots showing lack of convergence of two chains for LL from IDR model fit to ILI from the 15-44 age group.**

The pooled to within chains variability (top (red) line in figures D.3 and D.4) is above 1 from the start of the simulation for all LOGR fits except fits to P&I data for the 5-14 and $\geq$ 65 age groups and to ILI for the

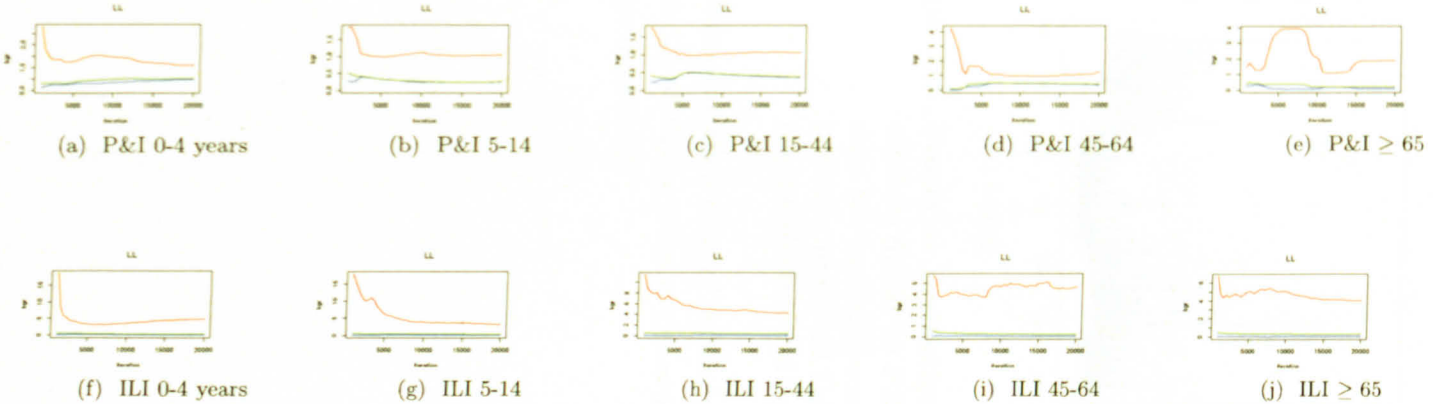
45-64 age group (figure D.3). This shows that for all but these 3 model fits, initial values for two chains of LOGR model were sufficiently disparate. The pooled to within chains variability is above 1 for all IDR mode fits showing initial values were suitably disparate for all IDR model fits (figure D.4).



(a) P&I 0-4 years (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64 (e) P&I $\geq$ 65

(f) ILI 0-4 years (g) ILI 5-14 (h) ILI 15-44 (i) ILI 45-64 (j) ILI $\geq$ 65

Figure D.3.: **Brooks-Gelman-Rubin plots of the LL from LOGR models after 20,000 iterations.** The top (red) line is the ratio of pooled to within chains variability; the two lines along the bottom are the width of the central 80% interval of the pooled runs (green) and the average width of the 80% intervals within individual runs (blue). Brooks-Gelman-Rubin plots monitor convergence of the ratio of pooled to within chains variability to 1 and both the width of the central 80% interval of the pooled runs and the average width of the 80% intervals within the individual runs to stability.

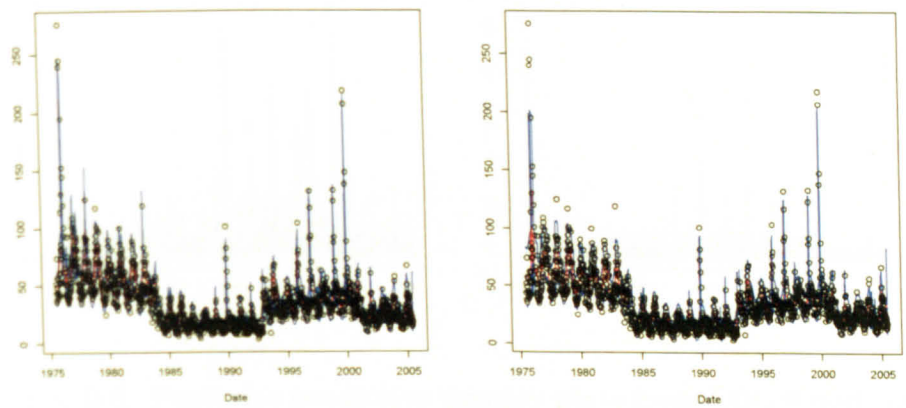


(a) P&I 0-4 years (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64 (e) P&I $\geq$ 65

(f) ILI 0-4 years (g) ILI 5-14 (h) ILI 15-44 (i) ILI 45-64 (j) ILI $\geq$ 65

Figure D.4.: **Brooks-Gelman-Rubin plots of the LL from IDR models after 20,000 iterations.**

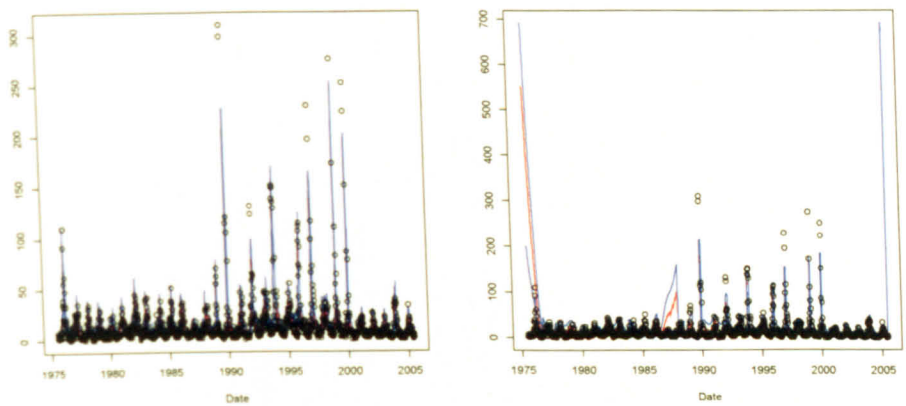
Figure D.5 shows better fit of the IDR than LOGR model to P&I from the 45-64 age group.



(a) LOGR (b) IDR

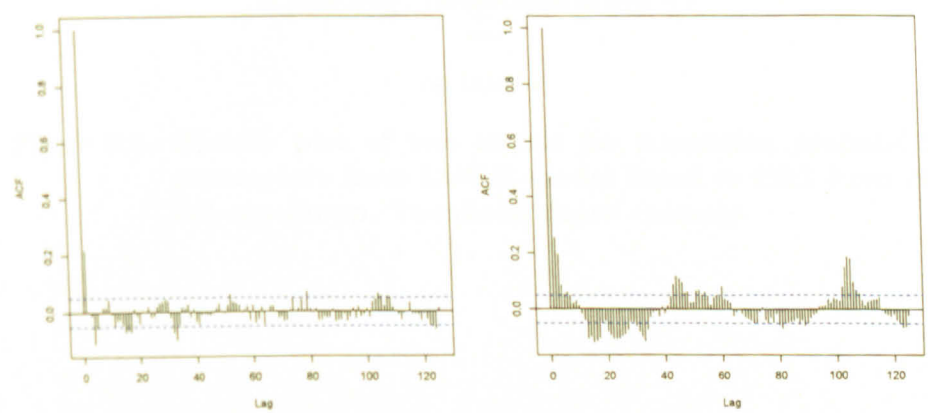
Figure D.5.: **Posterior predictive density plots of LOGR and IDR models fitted to P&I from the 45-64 age group.** These data are underdispersed (almost no observed data outside the posterior predicted CrIs) relative to the LOGR model (a) but adequately modeled by the IDR model (b).

Autocorrelation plots of residuals from IDR and LOGR model fit to P&I from the 45-64 age group are shown below (figure D.7). The IDR model fits these data badly.



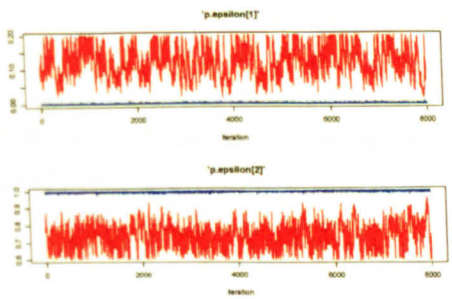
(a) LOGR (b) IDR

Figure D.6.: **Posterior predictive density plots from LOGR and IDR models fitted to ILI from the ≥ 65 age group.** The LOGR model (a) adequately models these data; note large posterior predictive CrIs at start and end of the time series for the IDR fit (b), indicating poor fit to these influenza seasons.



(a) LOGR (b) IDR

Figure D.7.: **Autocorrelation plots of residuals from LOGR and IDR model fits to P&I from the 45-64 age group.** The correlation between residuals against the lag between the residuals. Correlation between residuals is 1 at lag 0 because this correlation is between the residual and itself. Horizontal dotted lines are set at $\pm 2/\sqrt{1566}$. These data are poorly modeled by IDR (b) and adequately modeled by LOGR (a).



(a) LOGR

Figure D.8.: **History plot of two chains for transition probability parameters from LOGR model fitted to P&I from the 0-4 age group.** Two chains do not converge.

E. Multivariate models

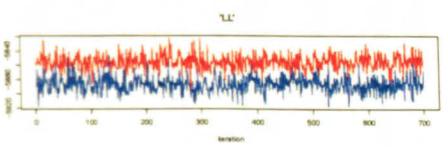
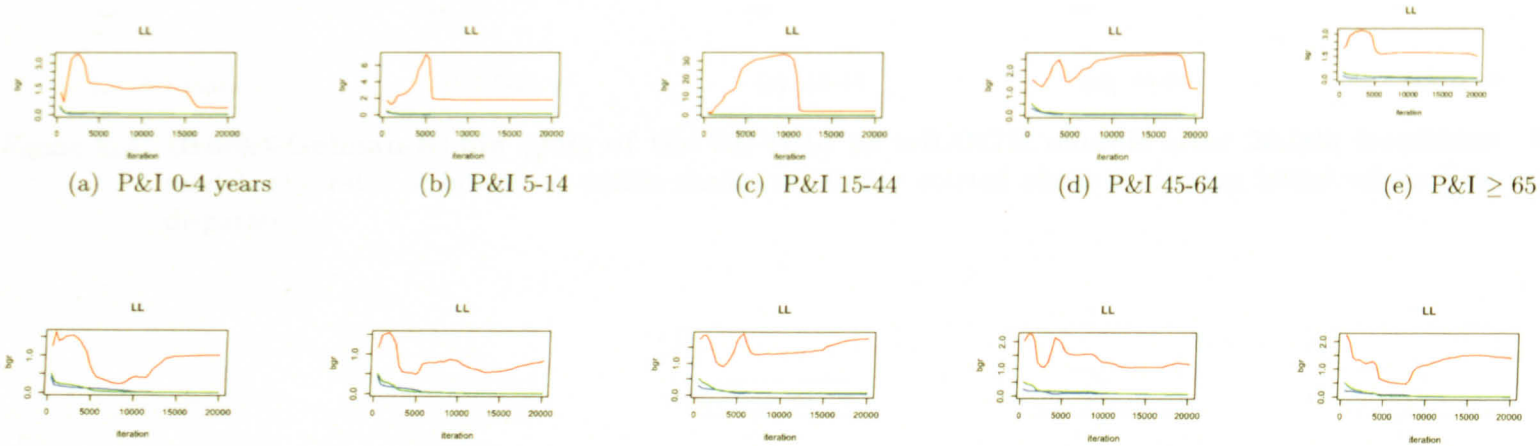


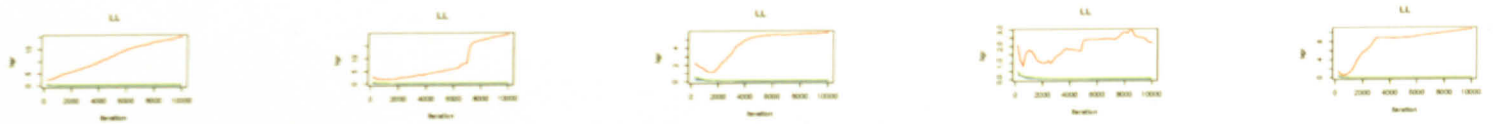
Figure E.1.: History plot showing lack of convergence of LL from biLOGR fit to P&I and laboratory reports from the 0-4 age group.



(a) P&I 0-4 years (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64 (e) P&I $\geq$ 65

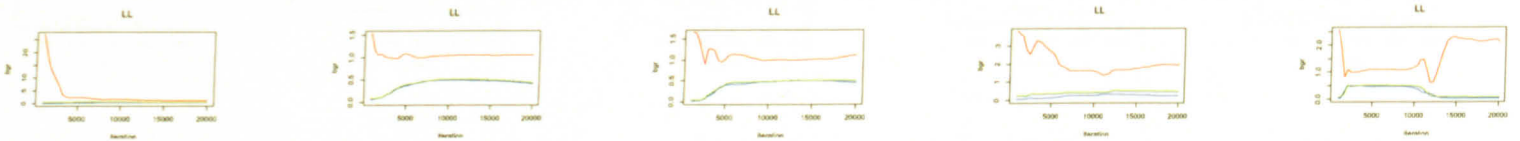
(f) ILI 0-4 years (g) ILI 5-14 (h) ILI 15-44 (i) ILI 45-64 (j) ILI $\geq$ 65

Figure E.2.: **Brooks-Gelman-Rubin plots of the LL (LL) of biLOGR models after 20,000 iterations.** The top (red) line is the ratio of pooled to within chains variability. Brooks-Gelman-Rubin plots monitor convergence of this ratio to 1. For all models, except the fit to P&I in the 15-44 age group, the ratio of pooled to within chains variability started above 1, showing initial values were suitably disparate. The two lines along the bottom are the width of the central 80% interval of the pooled runs (green) and the average width of the 80% intervals within individual runs (blue). These metrics should converge to stability.

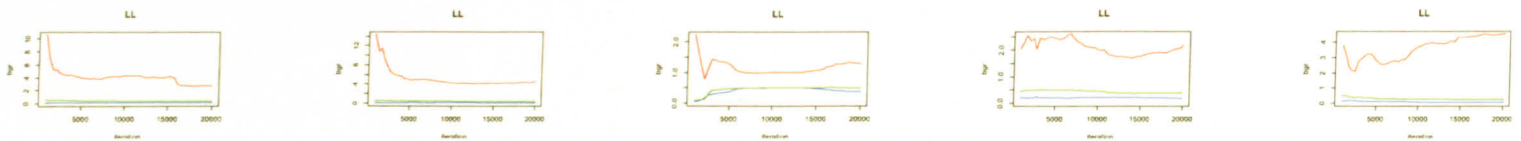


(a) 0-4 years (b) 5-14 (c) 15-44 (d) 45-64 (e) ≥ 65

Figure E.3.: **Brooks-Gelman-Rubin plots of the LL (LL) of triLOGR models after 20,000 iterations.** For all models the ratio of pooled to within chains variability started above 1 showing initial values were suitably disparate.

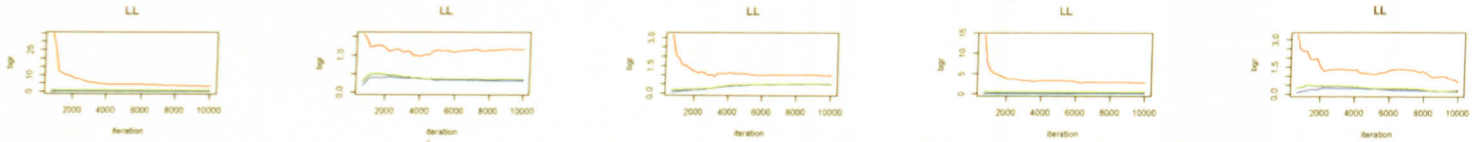


(a) P&I 0-4 years (b) P&I 5-14 (c) P&I 15-44 (d) P&I 45-64 (e) P&I ≥ 65



(f) ILI 0-4 years (g) ILI 5-14 (h) ILI 15-44 (i) ILI 45-64 (j) ILI ≥ 65

Figure E.4.: **Brooks-Gelman-Rubin plots of the LL (LL) of biIDR models after 20,000 iterations.** Initial values were suitably disparate for all models.



(a) 0-4 years (b) 5-14 (c) 15-44 (d) 45-64 (e) ≥ 65

Figure E.5.: **Brooks-Gelman-Rubin plots of the LL (LL) of triIDR models after 20,000 iterations.** Initial values were suitably disparate for all models.

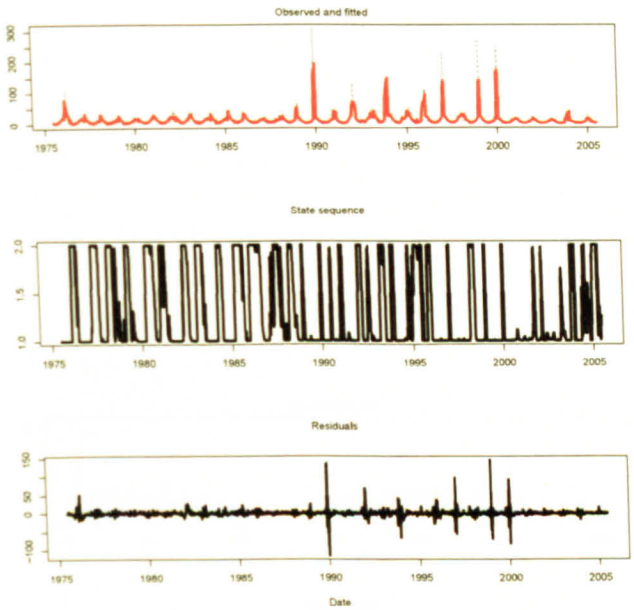


Figure E.6.: The state-sequence for biLOGR model fit to ILI and laboratory reports from the ≥ 65 age group (shown) appears to converge for all but the 2002/03 season, during which there is disagreement between the two chains as to the state assignment.

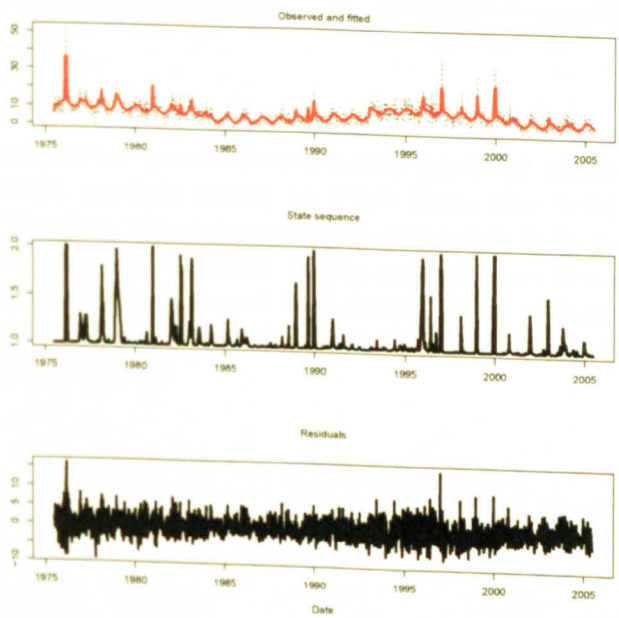
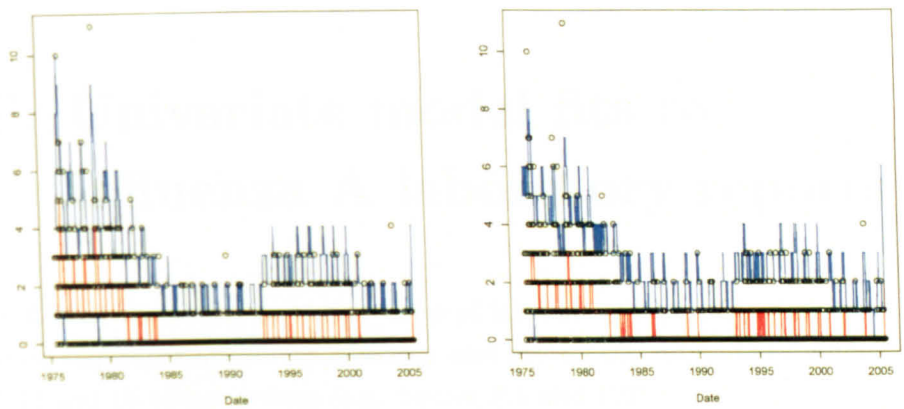


Figure E.7.: State-sequences for IDR models fitted to P&I in those 15-44 years of age.



(a) biLOGR (b) biIDR

Figure E.8.: **Posterior predictive density plots of biLOGR and biIDR models fitted to P&I from the 5-14 age group.** The data are underdispersed relative to both models (almost all observed data fall within posterior predictive CrIs). The predicted counts (red line) and a CrI for each predicted count (blue lines) and the observed P&I data for the 5-14 age group (black circles).

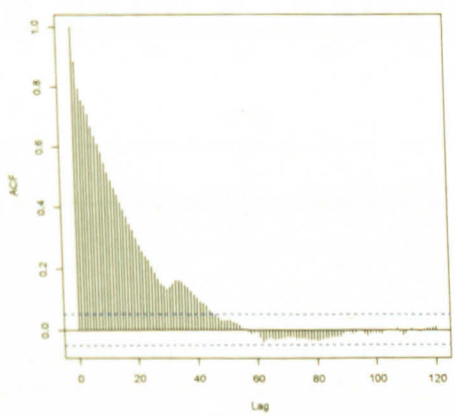


Figure E.9.: **Autocorrelation plot of residuals from biIDR model fit to ILI from the 5-14 age group.** The model's accounting for autocorrelation in the ILI data for the 5-14 age group was poor, with residual autocorrelation present at lags up to approximately 45 weeks. Horizontal dotted lines are set at $\pm 2/\sqrt{1566}$.

F. Univariate model fits to influenza A laboratory reports

In LOGR model fits to weekly rates of laboratory reports for influenza A, state sequences appear to converge and are clearly estimated for the 0-4, 5-14 and 15-44 age groups (e.g. figures F.1 and F.2).

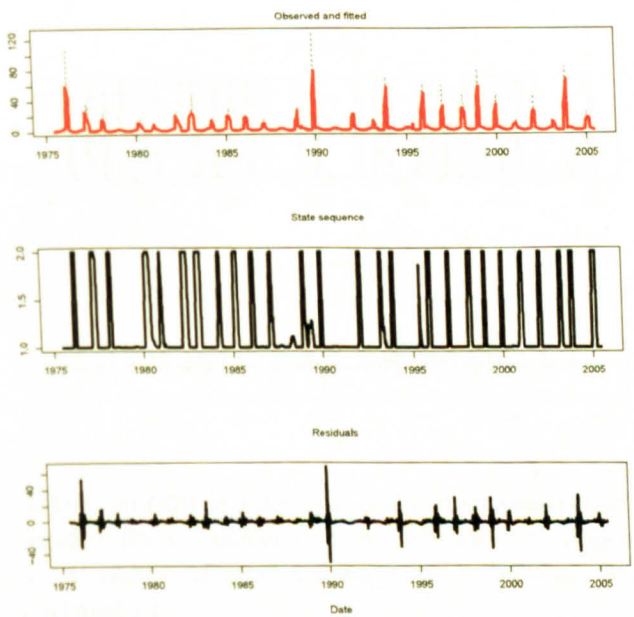


Figure F.1.: **After 20,000 iterations, the state-sequence for LOGR model fit to influenza A laboratory reports in those 0-4 years of age appears to have converged and is clearly estimated.**

In IDR model fits to weekly counts of influenza A laboratory reports, state sequences appear to converge and are clearly estimated for the 0-4 and 5-14 age groups (figures F.3 to F.4).

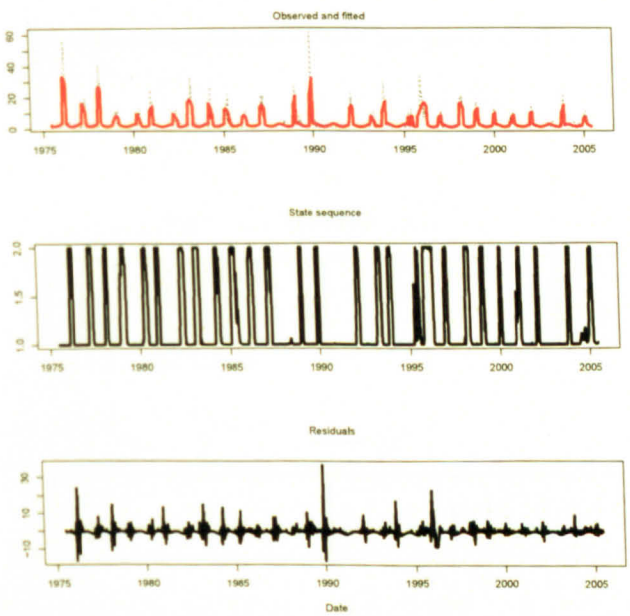


Figure F.2.: After 20,000 iterations, the state-sequence for LOGR model fit to influenza A laboratory reports in those 5-14 years of age appears to converge and is clearly estimated.

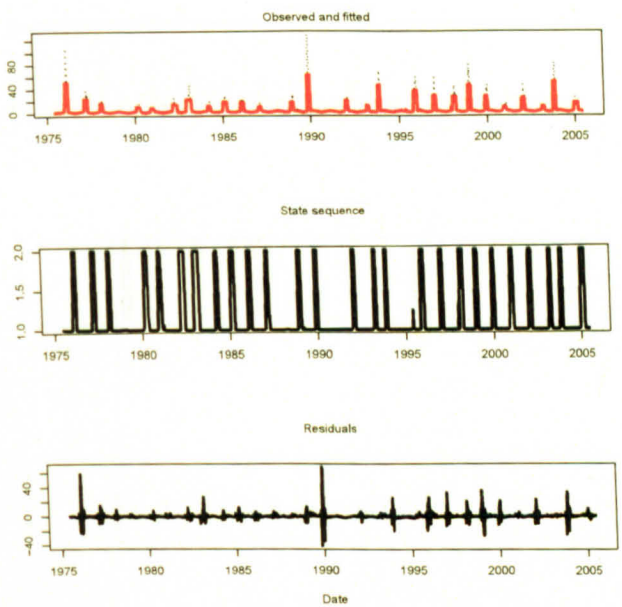


Figure F.3.: After 20,000 iterations, the state-sequence for IDR model fit to influenza A laboratory reports in those 0-4 years of age appears to converge and is clearly estimated.

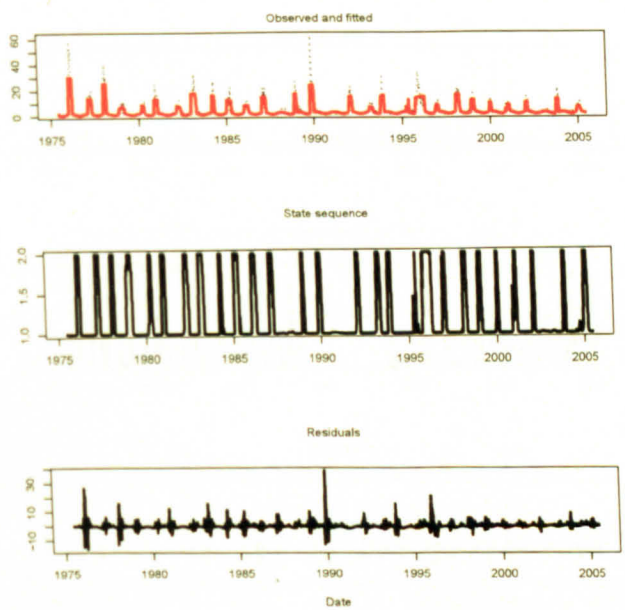


Figure F.4.: After 20,000 iterations, the state-sequence for IDR model fit to influenza A laboratory reports in those 5-14 years of age appears to converge and is clearly estimated.

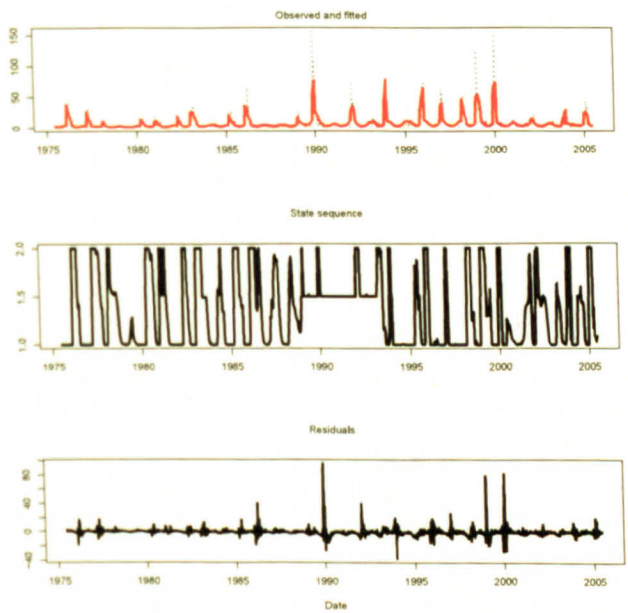


Figure F.5.: After 20,000 iterations, the state-sequence for LOGR model fit to influenza A laboratory reports in those $\geq$ 65 has not converged in the period 1988/89 to 1993/94.

G. Evidence for lag

Evidence for state transitions for a particular influenza season occurring in different weeks for different outcome variables is seen by overlaying age group-specific univariate ILI, P&I and laboratory report model state sequences (e.g. figures G.1 and G.2).

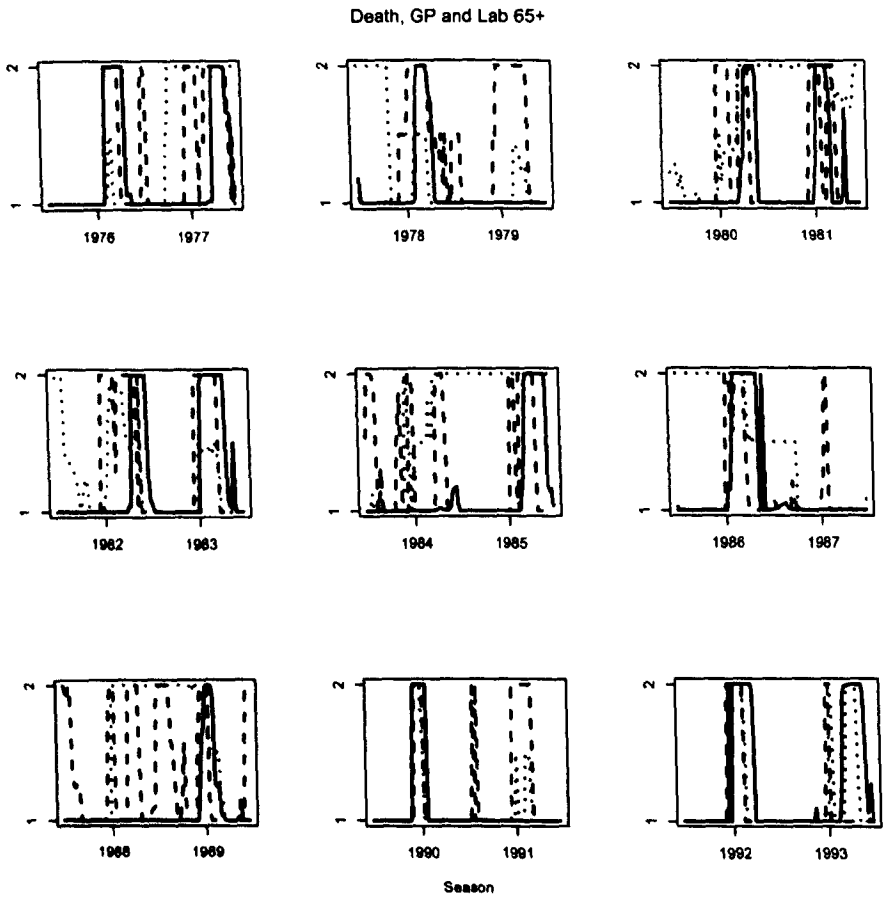


Figure G.1.: **State sequence from IDR fits to P&I (dashed line), ILI (dotted line) and lab reports for influenza A (solid line) from the $\geq$ 65 age group: 1975/76-1992/93.** The state sequence for the ILI model fit did not converge for all seasons. A time-varying lag between state transitions for one outcome and for another is evident.

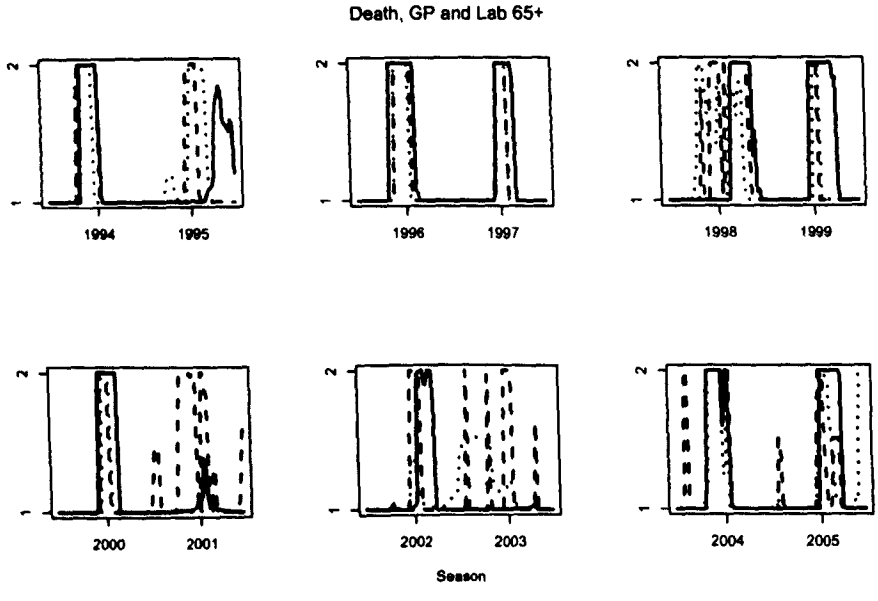


Figure G.2.: **State sequence from IDR fits to P&I (dashed line), ILI (dotted line) and lab reports for influenza A (solid line) from the $\geq$ 65 age group: 1993/04-2004/05.** The state sequences do not converge for all seasons. A lag between state transitions across outcomes is evident and it varies by season.

The formula below is for a model with a lagged state transition effect on P&I relative to its effect on laboratory reports.

$$
\begin{aligned}
Y_{P\&I_t} &\sim \text{Poisson}(\mu_{P\&I_t}) \\
Y_{lab_t} &\sim \text{Poisson}(\mu_{lab_t})
\end{aligned}
$$

$$
\begin{aligned}
\log(\mu_{P\&I_t}) |\, S_{t_{lagged}} = 1 &= \log(N_{P\&I_t}) + \alpha_{0_{P\&I}} + C_{P\&I}(t, \varphi_{P\&I}) \\
&+ \beta_{1_{P\&I}} \sin \tfrac{2\pi t}{52.2} + \beta_{2_{P\&I}} \cos \tfrac{2\pi t}{52.2} \\
&+ \beta_{3_{P\&I}} \text{artifacts}
\end{aligned}
$$

$$
\begin{aligned}
\log(\mu_{P\&I_t}) |\, S_{t_{lagged}} = 2 &= \log(N_{P\&I_t}) + \alpha_{0_{P\&I}} + \alpha_{P\&I}[\text{flu season}] + C_{P\&I}(t, \varphi_{P\&I}) \\
&+ \beta_{1_{P\&I}} \sin \tfrac{2\pi t}{52.2} + \beta_{2_{P\&I}} \cos \tfrac{2\pi t}{52.2} \\
&+ \beta_{3_{P\&I}} \text{artifacts}
\end{aligned}
$$

$$
\begin{aligned}
\log(\mu_{lab_t}) |\, S_t = 1 &= \log(N_{lab_t}) + \alpha_{0_{lab}} + C_{lab}(t, \varphi_{lab}) \\
&+ \beta_{1_{lab}} \sin \tfrac{2\pi t}{52.2} + \beta_{2_{lab}} \cos \tfrac{2\pi t}{52.2} \\
&+ \beta_{3_{lab}} \text{artifacts}
\end{aligned}
$$

$$
\begin{aligned}
\log(\mu_{lab_t}) |\, S_t = 2 &= \log(N_{lab_t}) + \alpha_{0_{lab}} + \alpha_{lab}[\text{flu season}] + C_{lab}(t, \varphi_{lab}) \\
&+ \beta_{1_{lab}} \sin \tfrac{2\pi t}{52.2} + \beta_{2_{lab}} \cos \tfrac{2\pi t}{52.2} \\
&+ \beta_{3_{lab}} \text{artifacts}
\end{aligned}
$$

$$
S_t |\, S_{t-1} \sim \text{Bernoulli}(\delta)
$$

$$
t_{lagged} = t - lag + 3
$$

$$
lag \sim \text{dcat}(p[1{:}5])
$$

Where Y_t are the observed number of P&I deaths or laboratory-confirmed influenza A cases in week t,
μ_t are the respective means of the Poisson distributions from which Y_t are drawn,
N_t are the populations at risk in week t,
α_0 are the intercepts,
α[flu season] are the yearly random effect mean shifts,
$C(t, \varphi)$ are the cubic splines with φ df, respectively,
$\beta_1 \sin \frac{2\pi t}{52.2} + \beta_2 \cos \frac{2\pi t}{52.2}$ represent seasonality,
$\beta_3$artifacts represent the instantaneous change in the baseline because of artifacts in the data,
and S_t is the state variable sampled from a Bernoulli distribution with probability δ, a two-by-two matrix of probabilities of moving (or not) between states at time t given which state the model was it at time t-1.

Each of the 5 possible lags ($\pm$ 2, 1 or 0 weeks) was assigned equal prior probability ($p = 0.20$).